



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학박사학위논문

흉선 세포 선별과정에서 E 단백질의 타겟 유전자 발현을  
억제하는 Twist2 의 기능에 관한 연구

**Study on the function of Twist2 in the regulation of  
E protein target gene expression during thymocyte selection**

2016 년 8 월

서울대학교 대학원

협동과정 유전공학전공

오 승 준

# **ABSTRACT**

## **Study on the function of Twist2 in the regulation of E protein target gene expression during thymocyte selection**

**Seungjoon Oh**

**Interdisciplinary program for Genetic Engineering**

**The Graduate School**

**Seoul National University**

Cellular differentiation is the process where a cell changes its state from one to another. The development of a multicellular organism involves a series of differentiation processes which result in a complex system of tissues and cell types. Differentiation is guided by the *de novo* binding of transcription factors to their cognate genomic sites. With the advent of technologies for investigating the genome-wide occupancy of transcription factors, a wealth of insights has emerged. Currently, ChIP-seq offers the most powerful means to locate genomic regions where critical regulatory interactions take place in a high resolution. Twist2, a bHLH transcription factor, is highly induced during thymocyte

selection in response to positively selecting TCR signals. Despite the importance of the positive selection-specific regulation of Twist2 expression for the proper differentiation of T lymphocytes, our understanding has been limited. In an effort to uncover the role of Twist2 during thymocyte selection, I intensively investigated the identities of the genes regulated by Twist2, and here I report the first genome-wide information about Twist2 occupancy in mouse thymocytes.

By performing ChIP-seq with thymocytes from the Twist2-transgenic mice, a total of 3,822 highly confident Twist2-binding sites were identified. The gene-proximal distribution of Twist2-binding sites was revealed. In addition, strong evolutionary conservation over Twist2-binding sites was exhibited, suggesting that these regions could be under selective pressure and therefore might be functionally important. By using *de novo* motif-discovery algorithm, the canonical E-box sequence (5'-CABMTGB-3') and two additional consensus sequences for the binding of ETS (5'-MGGAAR-3') family and Runx (5'-TGTGGTT-3') family transcription factors were discovered. E-box is the consensus DNA sequence to which bHLH transcription factors bind. In general, bHLH proteins dimerize preferentially with other members of the bHLH family to form hetero dimers. Before thymocyte selection, E proteins (HEB and E2A) are highly expressed in DP thymocytes and regulate the expression of critical genes by binding to E-box sequences located in the transcriptional regulatory regions.

I hypothesized that Twist2 may bind to the E-box sites pre-occupied by E proteins after being induced in post-DP thymocytes. To validate this tentative hypothesis, a recent study performed by Vanhille et al., which identified HEB-binding sites in DP thymocytes, was exploited and binding profiles of HEB and Twist2 were compared to count the number of regions shared by HEB and Twist2. Strikingly, 51% (842) of HEB-binding sites were overlapped with Twist2-binding sites. These overlapping genomic sites correspond to 22% (847/3,822) of Twist2-binding sites.

To identify putative genes directly regulated by Twist2 and/or E proteins in DP thymocytes, the occupancy peaks were linked to their adjacent genes. As results, 1,051 and 430 genes were associated with Twist2 and HEB, respectively. Among them, 239 genes were associated with the composite sites, binding of both Twist2 and HEB. To evaluate how the Twist2 occupancy correlates with the differential gene expression in DP thymocytes before (low level expression of Twist2) and immediately after (high level expression of Twist2) thymocyte selection, I analyzed the global gene expression profiles of pre-DP and post-DP thymocytes using microarray. The functional consequences of Twist2 binding to the composite sites during thymocyte selection were then estimated by Gene Set Enrichment Analysis (GSEA). As expected, genes associated with the composite sites were highly enriched among the group of genes down regulated upon thymocyte selection. Furthermore, genes associated with the composite sites were also highly enriched among the group of genes down regulated upon E protein deletion. This

indicated that E-protein activity is indeed crucial for the proper expression of genes associated with the composite sites in pre-DP thymocytes. These analyses strongly supported that Twist2 might negatively regulate the E protein target genes during thymocyte selection by binding to the genomic regions preoccupied by E proteins.

By using DO11.10 TCR-transgenic mouse system, effects of Twist2 knock out on the expression of 23 genes were analyzed. In the absence of Twist2, genes that are known to be down regulated in response to thymocyte selection were up regulated. For example, *Rorc* and *Cldn4*, whose expression dropped dramatically upon receipt of positive selection signal, were up regulated in Twist2<sup>f/f</sup>CD4Cre<sup>+</sup> DO11.10 cells. In contrast, *Poll* and *Trim26*, whose expressions were highly induced in positively selected DP cells, were down regulated in Twist2<sup>f/f</sup>CD4Cre<sup>+</sup> DO11.10 cells.

Here, I suggest a novel mechanism of the transcriptional repression of E protein-mediated pre-selection programs in DP thymocytes. After being highly induced in response to positively selecting TCR signals, Twist2, forms heterodimer with E proteins and thereby negatively regulates the expression of E protein target genes, resulting transition to post-selection stage.

*Keywords:* Twist2, ChIP-seq, thymocytes, development, bHLH

*Student Number:* 2008-30808

## TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>I</b>
<b>TABLE OF CONTENTS .....</b>	<b>V</b>
<b>LIST OF FIGURES .....</b>	<b>VII</b>
<b>LIST OF TABLES .....</b>	<b>X</b>
<b>INTRODUCTION .....</b>	<b>1</b>
I.    Cell-fate decision .....	1
II.   T lymphocytes .....	1
III.  T-cell development in the thymus .....	2
IV.   TCR signaling in thymocytes .....	8
V.    TCR-mediated positive selection of DP thymocytes .....	12
VI.   E protein transcription factors .....	13
VII.  Twist2, a member of the bHLH family of transcription factors .....	23
<b>MATERIALS AND METHOD .....</b>	<b>29</b>
<b>RESULTS .....</b>	<b>33</b>
<i>Genome-wide identification of in vivo Twist2-binding sites by ChIP-seq .....</i>	<i>33</i>

<i>Twist2 and HEB bind to overlapping sites on the DP-thymocytes genome</i> .....	52
<i>Twist2 negatively regulates HEB-target genes during thymocyte selection</i> .....	57
<i>Negative regulation of Rorc expression by Twist2 during thymocyte selection</i> ....	77
<b>DISCUSSION</b> .....	<b>92</b>
<i>Genome-wide Twist2-binding sites were identified by ChIP-seq</i> .....	93
<i>E-box motifs were most significantly enriched in Twist2-binding sites</i> .....	94
<i>Twist2 bind to overlapping sites with HEB on the DP-thymocytes genome and negatively regulates HEB target genes during thymocyte selection</i> .....	96
<i>Role of Twist2 in regulation of gene expression during thymocyte selection</i> .....	97
<i>Synergism between Twist2 and Id3 for the negative regulation of E protein target gene expression</i> .....	99
<i>Conclusion</i> .....	101
<b>REFERENCES</b> .....	<b>103</b>
<b>ABSTRACT IN KOREAN</b> .....	<b>118</b>
<b>ACKNOWLEDGEMENT</b> .....	<b>120</b>



## LIST OF FIGURES

Figure 1. T-cell development in the thymus .....	4
Figure 2. Generation of T-cell receptor (TCR) diversity by somatic recombination of TCR gene segments .....	6
Figure 3. TCR signaling in thymocytes .....	10
Figure 4. Structure and function of E proteins and Id proteins .....	15
Figure 5. Inhibition of E-protein activity by TCR signaling .....	18
Figure 6. Maintenance of the phenotype of DN3 and DP cells and enforcement of $\beta$ -selection and thymocyte-selection checkpoints by E proteins .....	21
Figure 7. Amino acid sequence alignment between Twist1 and Twist2 .....	26
Figure 8. Evaluation of the Twist2 ChIP-seq results .....	40
Figure 9. Genome-wide distribution of Twist2-binding sites relative to known and predicted genes in the mouse genome .....	43
Figure 10. Enrichment of Twist2-binding sites in defined genomic regions .....	45
Figure 11. Distance of Twist2-binding sites from the TSS of annotated genes .....	47

Figure 12. Evolutionary conservation of Twist2-binding sites .....	49
Figure 13. Consensus DNA sequences associated with Twist2 occupancy .....	51
Figure 14. Relative distribution of motifs across Twist2-binding sites .....	54
Figure 15. Common chromatin occupancy of Twist2 and HEB .....	59
Figure 16. Extent of overlapping between Twist2- and HEB-binding sites .....	61
Figure 17. Common target genes of Twist2 and HEB .....	64
Figure 18. Purification of cells for mRNA microarray analysis .....	66
Figure 19. Differential gene expression between pre- and post-DP thymocytes .....	68
Figure 20. Correlation between Twist2 occupancy and the gene expression changes upon thymocyte selection .....	71
Figure 21. Correlation between Twist2 occupancy and the gene expression changes upon E protein removal .....	74
Figure 22. Significant changes in the expression of genes associated with the composite sites during thymocyte selection as well as upon E protein removal .....	76
Figure 23. Effects of Twist2 knock out on the expression of genes associated with the composite sites .....	79
Figure 24. Composite binding of Twist2 and HEB to the ROR $\gamma$ t TSS .....	82

Figure 25. Two highly conserved E-box sites within ROR $\gamma$ t promoter region .....	84
Figure 26. Negative regulation of ROR $\gamma$ t expression by Twist2 .....	87
Figure 27. Negative regulation of ROR $\gamma$ t promoter activity by Twist2 .....	89
Figure 28. Direct binding of Twist2 to ROR $\gamma$ t promoter .....	91
Figure 29. Model of the Twist2-mediated regulation of gene expression during thymocyte selection .....	102

## LIST OF TABLES

Table 1. Properties of sequenced libraries .....	35
Table 2. Properties of identified peaks (1) .....	38
Table 3. Properties of identified peaks (2) .....	56

# INTRODUCTION

## I. Cell-fate decision

Cell-fate decision is crucial for the development of multicellular organisms, because proper diversity of cell types has to be generated for conducting various physiological functions. Most cellular decisions are controlled by signals transmitted from the cell-surface receptors (Freeman, 2000). T cells are particularly intriguing in that they express unique surface receptors generated by somatic DNA rearrangement. These clonal receptors determine which T-cell precursors will successfully mature.

## II. T lymphocytes

$\alpha\beta$  T cells, the essential mediators of adaptive immunity, express  $\alpha\beta$  T-cell receptors (TCRs) and respond to antigens in the form of short peptides bound to major histocompatibility complex (MHC) molecules (Germain, 1994). Typically,  $\alpha\beta$  T cells are distinguished by exclusive surface expression of either CD8 or CD4 (Biddison et al., 1982; Swain, 1983). These CD4 and CD8 are referred to as co-receptors because they cooperatively recognize TCR ligands (Janeway, 1992). CD8 binds to a MHC molecule, but is specific to the class I MHC protein. On the contrary, CD4 binds specifically to the

class II MHC protein. CD8<sup>+</sup> T cells (cytotoxic T cells) activated by MHC class-I ligands clear intracellular infection by targeting infected cells for destruction. Whereas, CD4<sup>+</sup> T cells (helper or regulatory T cells) triggered by MHC class-II ligands orchestrate intercellular co-operation between immune cells to combat extracellular pathogens.

### **III. T-cell development in the thymus**

Maintaining proper number and diversity of T cells is essential for optimal host defense. Thus, the development of T cells in the thymus is tightly regulated to ensure that the resulting T-cell population is responsive to vast and diverse foreign antigens, but not overtly autoreactive. This feat is achieved through a variety of developmental checkpoints guiding the generation of TCR components and regulating the survival of cells only those expressing functional, self-tolerant TCRs.

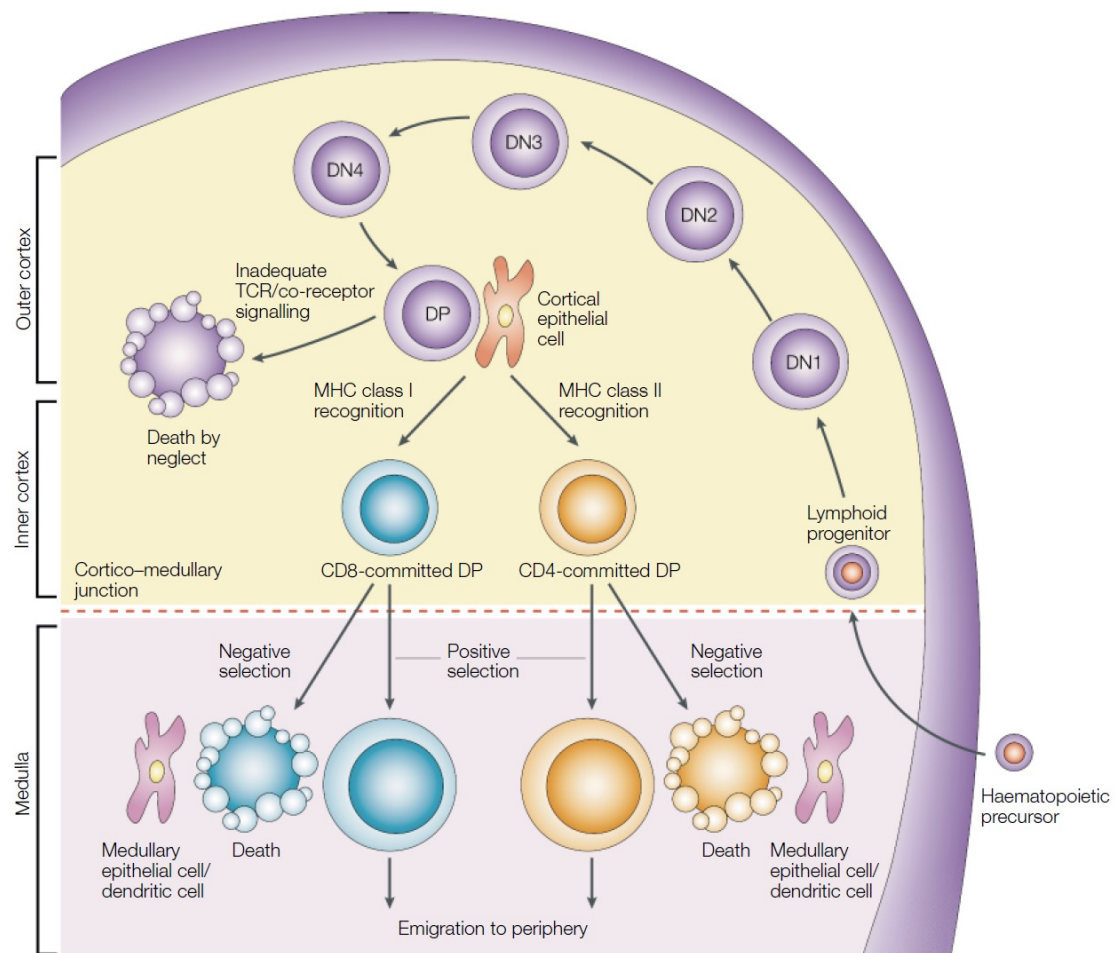
#### *Generation of CD4<sup>+</sup>CD8<sup>+</sup> thymocytes*

After being generated in the bone marrow, committed lymphoid progenitors migrate to the thymus through the blood. These cells lose potential for B-cell and natural-killer-cell development in the thymus (Radtke et al., 2009; Pui et al., 1999; Wilson et al., 2001; Michie et al., 2000). They are the committed T-cell precursors, called double-

negative (DN; no CD4 or CD8) thymocytes. They can give rise to either  $\gamma\delta$  or  $\alpha\beta$  T cells (Robey and Fowlkes, 1994) (Figure 1).

DN thymocytes can further be subdivided into four consecutive stages of differentiation, which are distinguishable by their surface expression of CD44 and CD25: DN1, CD44+CD25-; DN2, CD44+CD25+; DN3, CD44-CD25+; and DN4, CD44-CD25- (Godfrey et al., 1993). Cells that proceed along the  $\alpha\beta$  TCR pathway first express pre-TCR- $\alpha$  at the DN3 stage (Groettrup et al., 1993; von Boehmer and Fehling, 1997; Aifantis et al., 1999). DN3 thymocytes express the pre-TCR (pre-TCR- $\alpha$ -chain and TCR- $\beta$  chain as a pair). TCR- $\beta$  chain is the product of a set of somatic DNA rearrangements in the genomic locus encoding TCR- $\beta$  chain. Every single DN3 thymocytes randomly rearranges variable (V), joining (J), and diversity (D) gene segments in the TCRb locus (Figure 2). By this means, unique amino acid sequence in the variable domain of TCR- $\beta$  chains, responsible for recognizing processed antigens, are generated. TCR- $\beta$  chain gene rearrangement is achieved through a process known as V(D)J recombination that is mediated by recombination-activating gene 1 (RAG1) and RAG2 (Mombaerts et al., 1992; Shinkai et al., 1993) proteins. At the cell surface, the pre-TCR is associated with a collection of proteins involved in proximal signal transduction (van Oers et al., 1995). Active signaling through the pre-TCR is required for further T-cell maturation. Successful pre-TCR expression leads to cell proliferation. This process is referred to as ' $\beta$ -selection'.

T cells that successfully pass through  $\beta$ -selection (late DN3 and DN4) undergo



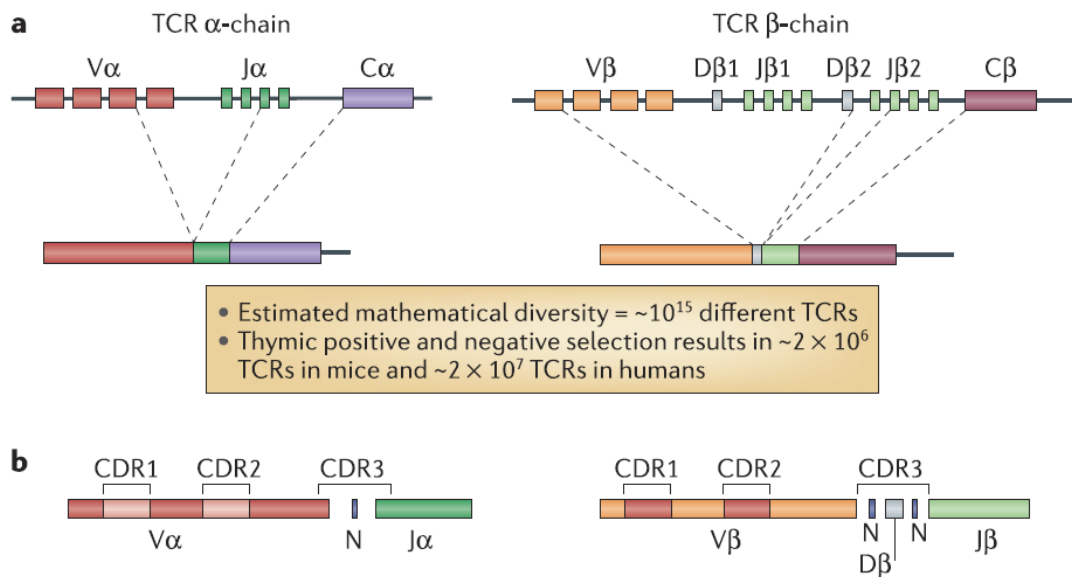
*Nature Rev. Immunol. 2002.*

**Figure 1. T-cell development in the thymus**



**Figure 2. Generation of T-cell receptor (TCR) diversity by somatic recombination of TCR gene segments**

(A) Functional TCRs are heterodimers consisting of an  $\alpha$ -chain and a  $\beta$ -chain. The TCR- $\beta$  chain is generated by somatic DNA recombination of variable (V), diversity (D) and junctional (J) gene segments. For the TCR- $\alpha$  chain, V and J gene segments are recombined. Rearranged gene segments are spliced together with the constant (C) region to form the functional  $\alpha\beta$  TCR. Each T cell express only one type of recombined receptor complex. (B) Hypervariability regions, known as complementarity-determining regions (CDRs), are encoded in the V gene segments (Chothia et al., 1988). Especially, CDR3 is created by the juxtaposition of different V(D)J germline segments after somatic recombination (Pannetier et al., 1993). The diversity of the TCR repertoire increased further by a lack of precision during V(D)J gene rearrangement and by the addition of non-template encoded nucleotides (N) at the V(D)J junctions (Cabaniols et al., 2001; Rock et al., 1994).



*Nature Rev. Immunol. 2006.*

several times of cell divisions. During the clonal expansion, the expression of pre-TCR- $\alpha$  is lost and the recombination at the TCRA locus takes place resulting in the production of the mature  $\alpha\beta$  TCR. In addition, these cells commence to express co-receptor proteins. Eventually, a large population of double-positive (DP; CD4+CD8+) thymocytes carrying clonotypic  $\alpha\beta$  TCR is formed. DP stage constitutes 90% of the thymocytes in the thymus of young individuals.

#### *TCR-mediated selection of DP thymocytes*

Among the large repertoire of DP thymocytes, the subset that is adequate for the host defense is allowed to mature and migrate to peripheral lymphoid organs. This process is known as 'Thymocyte selection' which can be subdivided into the four distinctive processes: death by neglect, negative selection, positive selection and lineage commitment (Robey and Fowlkes, 1994; von Boehmer et al., 1989).

Thymocyte selection is triggered by TCR signaling. And the ligand specificity of TCR determines the subsequent developmental fates of each thymocytes. Most DP thymocytes express TCRs that are incapable of recognizing the self-peptide in context of MHC molecules (self-pMHC). These cells are destined to die within 3-4 days since they cannot transmit intracellular signals that are necessary for maintaining viability (referred to as 'death by neglect'). A group of DP thymocytes bears TCRs that bind self-ligands

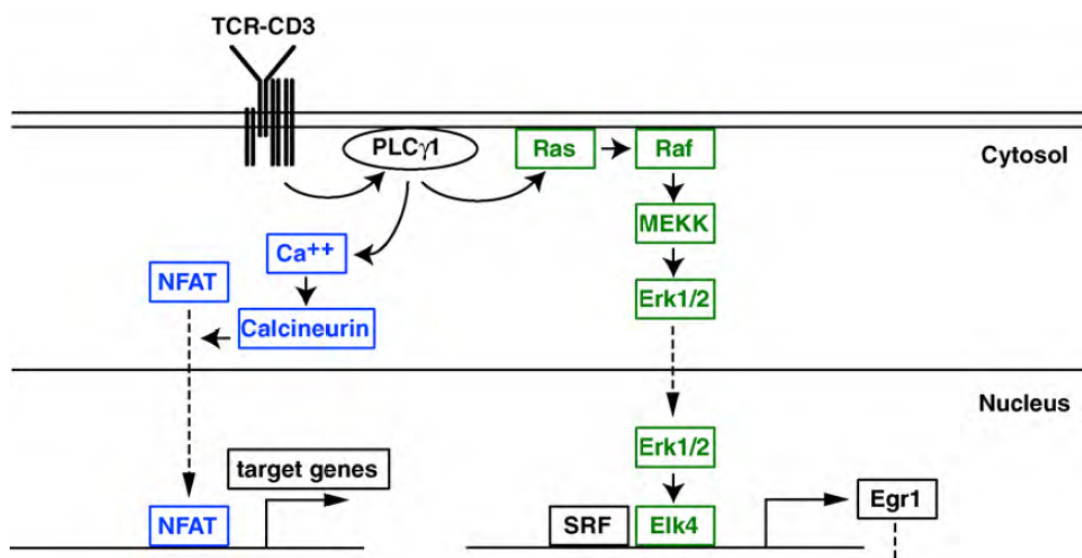
with high affinity. These lymphocytes could give rise to autoimmune diseases if they were permitted to emigrate from the thymus and populate peripheral lymphoid organs. Signaling on engagement of these autoreactive TCRs with self-pMHC transmits strong signals and promotes rapid apoptotic death (referred to as 'negative selection'). Consequently, only those cells expressing TCRs of moderate avidity for the self-pMHC escape programmed cell death, because their TCRs generate signals that have an intensity between those resulting in 'death by neglect' or 'negative selection'. These 'positively selected' thymocytes differentiate into the subsequent CD4<sup>+</sup> or CD8<sup>+</sup> (single-positive; SP) stage and complete maturation for helper or cytotoxic T cells. In summary, TCR-mediated selection of DP thymocytes results in a customized T-cell repertoire bearing potentially useful TCR.

#### **IV. TCR Signaling in thymocytes**

As TCR binds to self-pMHC on thymic epithelial cells, TCR complex transmits intracellular signal which diverge into the mitogen-activated protein kinase (MAPK) pathway and the Ca<sup>2+</sup>-Calcineurin pathway (Figure 3). Cytoplasmic tail of coreceptors are associated with protein tyrosine kinases, such as Lck. These protein tyrosine kinases phosphorylate immunoreceptor tyrosine-based activation motifs (ITAMs) on CD3 (Hashimoto et al., 1996; Sohn et al., 2001). ZAP-70, another protein tyrosine kinase, is

### **Figure 3. TCR signaling in thymocytes**

Schematic representation of the MAPK (green) and the  $\text{Ca}^{2+}$ -Calcineurin (blue) pathways in thymocytes. Plain arrows indicate activation and dashed arrows indicate nuclear import. Active  $\text{PLC}\gamma 1$  hydrolyses the membrane-associated  $\text{PIP}_2$ . The released  $\text{IP}_3$  and DAG causes calcium influx and activates Ras, respectively.



*Seminars in Immunol. 2010.*

recruited to the phosphorylated tyrosine residues on the CD3 and phosphorylated by Lck (Negishi et al., 1995). The activated ZAP-70 in turn phosphorylates an adaptor protein, linker for activation of T cells (LAT) (Zhang et al., 1999). Phospholipase C $\gamma$ 1 (PLC $\gamma$ 1) is then recruited to the cell membrane by binding to the phosphorylated LAT (Wange, 2000). PLC $\gamma$ 1 hydrolyses the membrane-associated phosphatidylinositol-3,4-bisphosphate (PIP<sub>2</sub>) into two secondary messenger molecules: diacylglycerol (DAG) and inositol-1,4,5-triphosphate (IP<sub>3</sub>). DAG and IP<sub>3</sub> are the critical inducers of the downstream MAPK and Ca<sup>2+</sup> signaling, respectively.

#### *MAPK pathway*

The MAPK pathway (also known as the Ras-Raf-MEK-ERK pathway) is triggered as DAG binds to RasGRP, a Ras guanine nucleotide exchange factor (RasGEF) (Dower et al., 2000). Via the MAPK pathway, various downstream transcription factors are phosphorylated and activated to regulate gene expression.

#### *Ca<sup>2+</sup> - Calcineurin pathway*

While DAG stays inside the membrane, IP<sub>3</sub> is soluble and diffuses through the cytoplasm. When IP<sub>3</sub> binds to IP<sub>3</sub> receptors on the membrane of endoplasmic reticulum

(ER),  $\text{Ca}^{2+}$  stored in the ER is released to the cytoplasm. The depletion of  $\text{Ca}^{2+}$  in the ER results in the opening of  $\text{Ca}^{2+}$  release activated  $\text{Ca}^{2+}$  channels (CRAC) on the plasma membrane which allows the entry of extracellular  $\text{Ca}^{2+}$  into the cytoplasm for the sustained influx of  $\text{Ca}^{2+}$  (Feske, 2007).

The increase in the cytoplasmic  $\text{Ca}^{2+}$  level results in the activation of Calcineurin, a  $\text{Ca}^{2+}$ -dependent serine/threonine phosphatase. As Calmodulin is associated with  $\text{Ca}^{2+}$ , it binds to the regulatory subunit of calcineurin (CnB1). And then, the autoinhibitory domain of Calcineurin is displaced from the active subunit of Calcineurin (CnAa and CnAb). The activated Calcineurin dephosphorylates nuclear factor of activated T cells (NFAT). By which, the nuclear localization sequence (NLS) in NFAT is unmasked, and NFAT proteins are able to translocate into the nucleus (Rao et al., 1997).

## **V. TCR-mediated positive selection of DP thymocytes**

The MAPK and the  $\text{Ca}^{2+}$  signaling pathways are required for the positive selection (Pagès et al., 1999). When either signaling pathway is disrupted, severe defects in the positive selection occurs. The RasGRP-deficient mice and the Erk1/2 conditional knockout mice display block in the positive selection (Fischer et al., 2005). The transgenic expression of dominant-negative or constitutively active forms of Ras, Raf or MEK1 exert effect on the positive selection, as well (Swan et al., 1995; Alberola-Ila, 1995; O'Shea et



al., 1996).

The  $\text{Ca}^{2+}$ -mediated activation of Calcineurin and NFATc3 is also implicated in the positive selection. Maturation of CD4 and CD8 single-positive thymocytes is prevented when cyclosporine A, which blocks the Calcineurin-dependent  $\text{Ca}^{2+}$  signaling, is treated (Gao et al., 1988; Jenkins et al., 1988). Calcineurin and NFATc3-knockout mice also show severe developmental defect in positive selection with no mature SP thymocytes (Oukka et al., 1998; Cante-Barrett et al., 2007; Neilson et al., 2004).

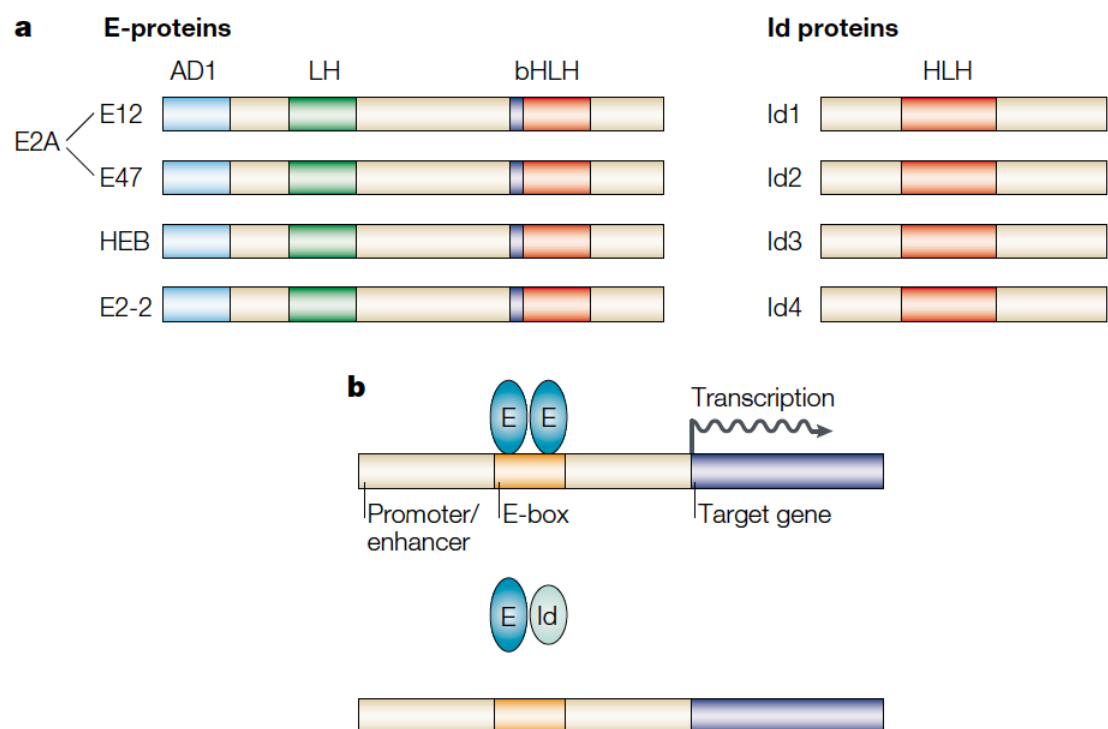
## **VI. E protein transcription factors**

E proteins, a family of basic helix-loop-helix (bHLH) transcription factors, have been reported to function as important transcriptional regulators in a large array of tissues and cell types, including neural, muscle, and lymphocytes of the immune system (Massari and Murre, 2000; Murre, 2005). They are implicated in basic cellular processes, such as cell differentiation, cell cycle, and survival.

E proteins consist of three members: E2A, HEB, and E2-2 (Figure 4). *Tcf2a* (also known as *Tcf3*) produces two alternative splicing variants, E12 and E47, which are collectively referred to as E2A (Murre et al., 1989). *Tcf12* and *Tcf4* encodes HEB and E2-2, respectively (Wang et al., 2006; Henthorn et al., 1990; Hu et al., 1992). The conserved bHLH domain is critical for the function of E proteins. The basic region mediates DNA

**Figure 4. Structure and function of E proteins and Id proteins**

(A) Schematic diagram of the E and Id proteins. The AD1- and LH(AD2)-activation domains, the basic (b) DNA-binding domains and helix–loop–helix (HLH) dimerization domains are highlighted. (B) Illustration of the binding of E-protein dimers to E-box sites located in promoters or enhancers and activation of the transcription of target genes. Since Id proteins lack the basic domain required for DNA binding, Id–E-protein dimers are thereby unable to bind to promoters or enhancers and activate transcription.



*Nature Rev. Immunol., 2001.*

binding to E-box sites (CANNTG) and the HLH domain mediates dimerization (Murre et al., 1989). E proteins function either as homodimers or as heterodimers with other E proteins or various bHLH factors to bind DNA and regulate gene expression (Massari and Murre, 2000).

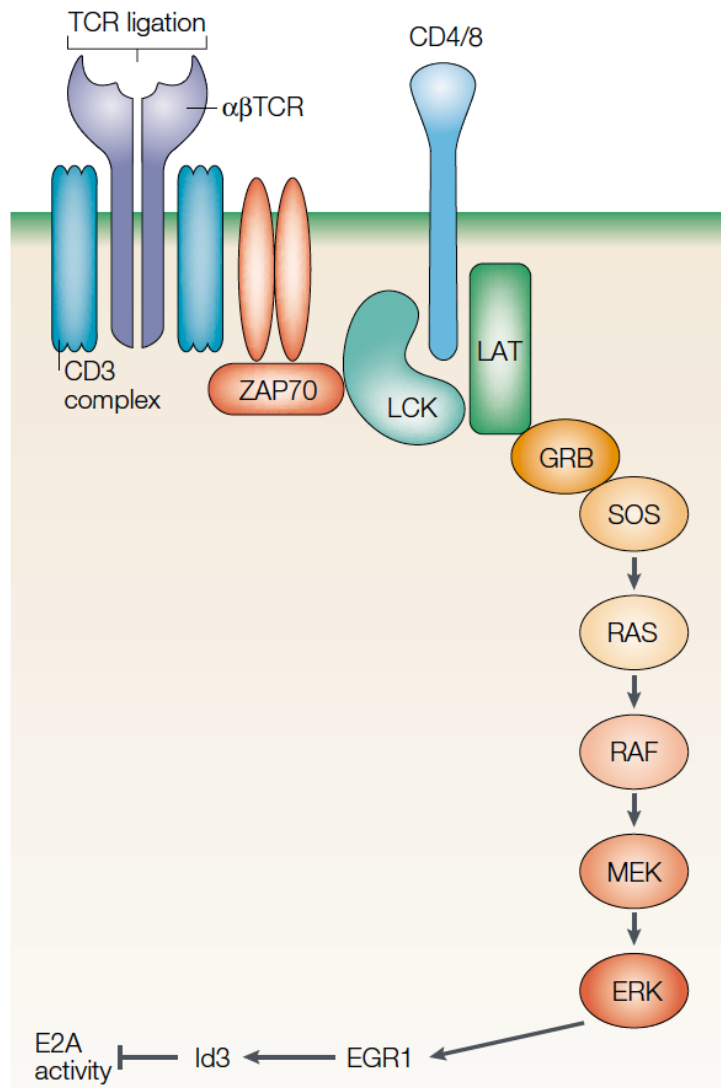
#### *TCR-mediated inhibition of E-protein activity*

While E proteins are broadly expressed in lymphoid-lineage cells, their expression level varies dynamically from stage to stage (Engel et al., 2001; David-Fung et al., 2006; Pan et al., 2002). Beyond regulation of E-protein abundance, the DNA-binding activity of E proteins is also tightly regulated in a stage-specific manner. Id protein family (Id1-4), another important group of factors that dimerize with E proteins, contain an HLH domain but lack the DNA-binding basic region. Id proteins therefore act as negative regulators of E protein since they can inhibit DNA binding by competitive dimerization with E proteins (Benezra et al., 1990). During particular stages of T-cell development, the expression of Id proteins is induced temporarily and the Id-mediated inhibition of the E-protein activity takes place (Engel et al., 2001; Bain et al., 2001; Koltsova et al., 2007; Rivera et al., 2000).

During thymocyte selection, TCR signals that initiate the positive selection induce Id3, which in turn inhibit E-protein activity (Figure 5). Evidence for such a

**Figure 5. Inhibition of E-protein activity by TCR signaling**

Schematic diagram depicting the inhibition of E2A activity by TCR-mediated activation of the MAP kinase pathway. (ERK, extracellular signal-regulated kinase; GRB, growth factor receptor bound; MAP, mitogen-activated protein kinase; MEK, MAPK kinase)



*Nature Rev. Immunol., 2001.*

mechanism has well been provided. *In vitro* crosslinking of TCR induces Id3 expression and inhibits the DNA-binding activity of E proteins in DP thymocytes (Bain et al., 2001). Moreover, both the induction of Id3 expression and the inhibition of E-protein activities are reduced by agents that block MEK1 and MEK2. They are the kinases involved in signal transduction through the MAPK pathway (Alessi et al., 1995; Dudley et al., 1995). Additional evidence indicates that early growth-response factor 1 (Egr1), the MAPK-responsive transcription factor, induces Id3 expression (Bain et al., 2001). Taken together, these data support a model in which TCR-mediated positive selection signal inhibits E-protein activity through Id3 that is induced by the activation of the MAP kinase cascade.

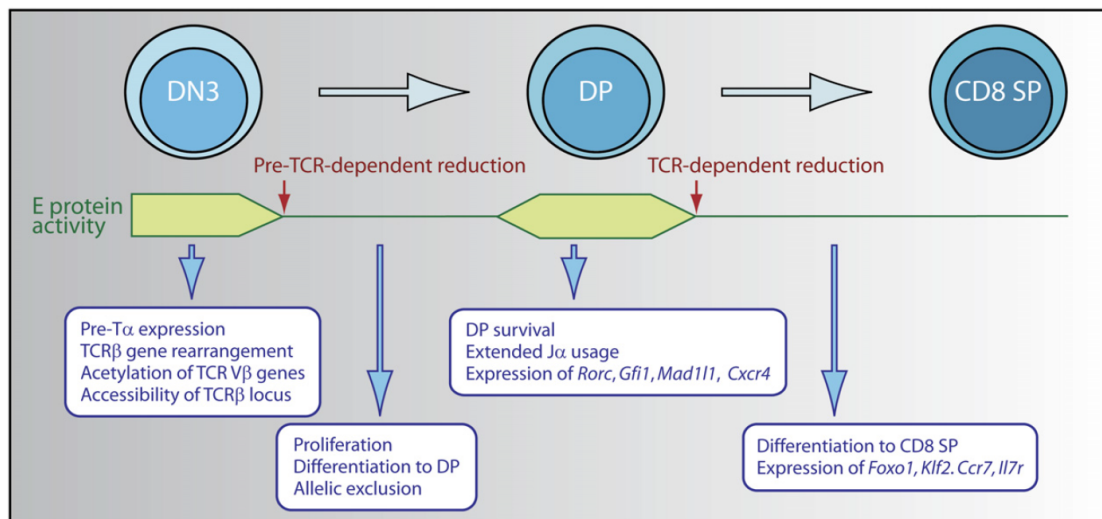
#### *Role of E proteins in T-cell development*

E proteins play critical roles throughout thymic development (Figure 6). E proteins regulate the maturation of early thymocytes. The E2A-deficient mice develop spontaneous T-cell leukemias at an early age, and these leukemic cells have the phenotype of immature thymocytes (Bain et al., 1997). E proteins also promote *Tcrg* and *Tcrd* gene rearrangement (Murre, 2005). In DN3 population, E proteins have multiple roles. E proteins maintain the phenotype of DN3 cells, enhance the expression of pre-T $\alpha$  and prevent proliferation (Engel and Murre, 2004). In addition, E proteins control the accessibility of the TCR $\beta$  locus (Agata et al., 2007). Pre-TCR signals reduce the activity

**Figure 6. Maintenance of the phenotype of DN3 and DP cells and enforcement of  $\beta$ -selection and thymocyte-selection checkpoints by E proteins**

DNA-binding activity of E proteins is high in DN3 and DP thymocytes. This activity is reduced in response to both pre-TCR and TCR signals. Some of the processes and genes that are controlled by E proteins at the indicated stages of thymocyte development are delineated.





*Immunity, 2007.*

as well as the expression of E proteins, and this allows differentiation of DN3 thymocytes to DP stage accompanied by several rounds of cell division. Thus, E proteins enforce the  $\beta$ -selection checkpoint.

In addition to the regulation of  $\beta$ -selection, E proteins are also implicated in controlling thymocyte selection. By analyzing the E2A and HEB conditional knockout thymocytes, Jones and Zhuang demonstrated that E proteins enforce the thymocyte-selection checkpoint (Jones and Zhuang, 2007). When E2A and HEB were deleted at the DP stage of thymocytes, the percentage of CD4-SP cells was decreased dramatically but the percentage of CD8-SP cells was increased. In the periphery, CD8-T cells which do not express the TCR are abundant. When several surface markers are examined, these TCR-negative CD8 cells are come out to have the phenotype of naïve resting T cells. These TCR-negative CD8+ T cells do not have functional TCR- $\alpha$  chain rearrangements but they make interferon- $\gamma$  in response to PMA and ionomycin stimulation. In conclusion, the lack of E-protein activity results in the differentiation of immature thymocytes to mature CD8-T cells regardless of functional TCR signals. These observations suggest the critical role of E proteins enforcing the positive-selection checkpoint in a manner similar to the way these proteins act on the  $\beta$ -selection checkpoint.

The authors also examined whether E proteins maintained the DP-thymocytes phenotype. With gene expression profiling of purified DP thymocytes, they found that genes known to be highly expressed in DP cells were down regulated by the deletion of

E2A and HEB. The microarray analysis also revealed that genes that are normally induced in response to positive selection were already elevated in pre-selection DP cells in the E2A- and HEB-deficient mice. These results solidify the role of E proteins in maintaining the DP-thymocytes phenotype until their activity is diminished by TCR signaling. In the absence of E proteins, the gene-expression program is altered, leading to differentiation into mature T cells.

As is the case for Id proteins, the presence or absence of binding partners alters the accessibility of E proteins to their target genes. Throughout developmental process, these factors act on E proteins to perform essential roles at the appropriate stage.

## **VII. Twist2, a member of the bHLH family of transcription factors**

Twist2 is a member of the bHLH family of transcription factors. The bHLH family transcription factors act as master regulators controlling cell fate in a variety of tissues from the early stages of embryogenesis. The bHLH transcription factors bind to DNA with a bipartite DNA-binding domain formed when two of these factors are associated as homo or heterodimer complexes through their HLH motifs. This DNA-binding domain recognize *cis*-regulatory elements containing the consensus sequence 5'-CANNTG-3' (termed E box). E boxes are widely distributed in the regulatory regions of numerous lineage-specifying genes (Wilson-Rawls et al., 2004; Murre et al., 1989;

Ellenberger et al., 1994).

The bHLH transcription factors are categorized into three major classes: the ubiquitous Class-A bHLH factors, E proteins (that include E2-2, HEB and the two isoforms of the E2A gene E12/E47); the tissue-restricted Class-B bHLH factors; and the inhibitory HLH proteins, Id proteins. The Twist proteins (Twist1 and Twist2) form a subfamily of the Class-B bHLH factors (Figure 7). Twist1 and Twist2 exhibit bifunctional roles as activators and repressors of gene transcription (Sharabi et al., 2008; Laursen et al., 2007).

Twist2 was first identified from a yeast-two-hybrid screen using E12 as bait, and it was named Dermo1 based on its expression pattern in the dermis of mouse embryos (Li et al., 1995). Dermo1 was later renamed as Twist2 according to its high degree of sequence similarity and overlapping expression pattern with Twist1. As seen with Twist1, Twist2 was found to inhibit both myogenic and osteoblast maturation (Li et al., 1995; Susic et al., 2003; Lee et al., 2000). During mammalian embryogenesis, Twist2 is expressed in mesodermal tissues. The Twist2-deficient mice exhibit normal embryonic development but typically dies around 2–3 days after birth due to cachexia, failure to thrive and high levels of pro-inflammatory cytokines (Susic et al., 2003).

Recently, it has been reported that two homozygous nonsense mutations (c.324 C>T and c.486 C>T) in TWIST2 cause Setleis Syndrome (MIM 227260) (Tukel et al.,

**Figure 7. Amino acid sequence alignment between Twist1 and Twist2**

Diagram showing the similarity between the two proteins. Black bars, Functional motifs; Red dots, Conserved threonine and serine phospho-regulated residues involved in dimer choice and DNA-binding site selection; Horizontal lines located above the alignment, Protein-protein interaction domains characterized for Twist1. The Twist box has been characterized as both an activation domain with the amino-acid motif LX<sub>3</sub>FX<sub>3</sub>R (indicated in red) and as a repressor domain.



2010). Setleis syndrome is an inherited developmental disorder classified as a Focal-Facial-Dermal-Dysplasia type III (FFDD III) and is characterized by bilateral temporal marks and additional facial features, including, absent eyelashes on both lids or multiple rows on the upper lids, absent Meibomian glands, slanted eyebrows and chin clefting. These mutations correspond to the glutamines 65 and 119 in the C-terminal domain of Twist2 protein.

The last 20 amino acids at the C-terminus contain a repressor domain termed 'Twist box' which is identical in both Twist1 and Twist2 (Bialek et al., 2004). A transactivation domain has also been characterized within the Twist box, containing amino-acid residues Leu-187, Phe-191 and Arg-195 in Twist1, which are completely conserved throughout the animal kingdom (Laursen et al., 2007). At the DNA level, Twist1 and Twist2 share a common intronic/exonic organization that reflects an evolutionary gene duplication event (Wolf et al., 1991; Tükel et al., 2010).

Twist2 has well been characterized as a master regulator of mesenchymal lineage development, encompassing myogenesis, osteogenesis and adipogenesis (Franco et al., 2011). In regard to hematopoiesis, Twist2 was first reported to play critical roles in the development of myeloid lineages (Sharabi et al., 2008). As evidenced by marked increase in mature myeloid populations of macrophages, neutrophils, and basophils in the Twist2-deficient mice, Sharabi et al. demonstrated the function of Twist2 as a negative regulator of proliferation and differentiation of myeloid-lineage progenitors. Recently, it was

identified that Twist2 expression is induced in response to TCR signaling during thymocyte selection (Oh, 2016). In particular, the expression of Twist2 is confined to the window of TCR-signal strength that results in the positive selection of thymocytes.

Despite the importance of the positive selection-specific regulation of gene expression for the proper differentiation of T lymphocytes, our understanding is limited. The expression of Twist2 is induced during thymocyte selection process, yet its detailed role remains unknown. In this study, I aimed to characterize the role of Twist2 during thymocyte selection by identifying its target genes. To elucidate target genes of Twist2, I performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) using Twist2-transgenic thymocytes, and multi-dimensional data analyses as well as further experimental validation.



## **MATERIALS AND METHODS**

### **Animals**

Twist2<sup>fl</sup> and Twist2-transgenic mice were previously described (Oh, 2016). CD4-Cre mice were purchased from The Jackson Laboratory. All mice were maintained on C57BL/6 background and bred at animal facility at Seoul National University. All animal experiments were performed in accordance with the guideline of Institutional Animal Care and Use Committees of Seoul National University.

### **Cell sorting**

Single-cell suspensions were prepared from thymus and stained with appropriate combinations of antibodies labeled with fluorochromes. The following antibodies were used: CD4-PE, CD8a-FITC, TCR $\beta$ -APC and CD5-PE-Cy7 (all from BD Biosciences). FACS Aria II (BD Biosciences) was used for cell sorting.

### **Twist2 Chromatin Immunoprecipitation (ChIP)**

ChIP was performed according to previously described methods (Lee et al.,

2006). Briefly,  $1 \times 10^7$  Twist2 transgenic whole thymocytes were fixed with 1% formaldehyde for 13 min at room temperature and cross-linking was stopped by the addition of glycine to a final concentration of 125mM. Sheared chromatin extract was prepared using Bioruptor UCD-200 (Diagenode). Dynal protein G beads (Invitrogen) coupled with appropriate antibodies were incubated with the chromatin extracts. Anti-Twist2 (H00117581-M01, Abnova) antibody was used for immunoprecipitation and rabbit IgG (sc-2027, Santa Cruz Biotechnology) was used as control. Precipitated immune-complex was reverse-crosslinked by heating and treated with Proteinase K. DNA was purified using QIAquick PCR Purification Kit (Qiagen). Input controls were prepared in parallel, but without immunoprecipitation step. The resulting ChIP-enriched DNA samples were analyzed using Prism 7900HT sequence detection system (Applied Biosystems).

### **ChIP-seq and bioinformatics analysis**

Twist2 ChIP-seq and Input DNA libraries were constructed with SOLiD ChIP-seq sample prep kit (Life Technologies Inc) according to the manufacturer's instructions. The libraries were sequenced on SOLiD 3.5 (Life Technologies Inc) with 50bp single-end reads. Twist2 ChIP-seq and Input reads were aligned to the mouse reference genome (mm9) using Bfast (0.7.0a) (Homer et al., 2009) with default parameters. Twist2-binding

sites were detected using Model-based Analysis for ChIP-Seq (MACS2, v2.1.0.20150603) (Zhang et al., 2008) with default parameter. The Twist2-binding sites were selected with previously published DNaseI ChIP-seq data (Vanhille et al., 2015). A PhastCons plot was generated using conservation-plot module on Cistrome Analysis Pipeline (Liu et al., 2011). Motif analysis was performed using MEME-ChIP suits (version 4.10.1) (Machanick et al., 2011) with UniPROBE mouse database. Peak annotation was carried out using GREAT (version 3.0.0) (McLean et al., 2010). Binding peaks were visualized with bedgraph file format using IGV (Robinson et al., 2011).

### **Integrating analysis with HEB ChIP-seq data**

Previously published HEB and Input ChIP-seq data (Vanhille et al., 2015; Lepoivre et al., 2013) were obtained from SRA (SRP050373, SRP019079). The SRA files were converted to Raw reads file (FASTQ) format with fastq-dump program from SRA toolkit. The resulting FASTQ files were aligned to the mouse reference genome (mm9) using Bfast (0.7.0a). HEB-binding sites were identified using MACS2 (MACS2, v2.1.0.20150603) with default parameters. Candidates of HEB-binding sites were selected with sharing DNaseI ChIP-seq data. The binding sites were annotated using GREAT (mm9).

## **RNA preparation**

Total RNA was isolated from FACS-purified thymocytes with TRIzol (Invitrogen) reagent, following the manufacturer's instructions. RNA quality was ensured using Experion (BioRad). Samples that had over 1.8 18S/28S rRNA ratios were selected for gene-expression profiling. Reverse transcription was carried out using a cDNA reverse transcription kit (Quantitect; Qiagen). SYBR Green PCR Master Mix and StepOnePlus sequence detection system (Applied Biosystem) were used for real-time quantitative PCR.

## **Gene-expression analysis using microarray**

Biotinylated RNA was prepared using the Illumina TotalPrep RNA Amplification Kit (Ambion) and hybridized to Illumina Mouse Ref-8 Expression BeadChip with quadruplicate for each samples. Hybridization signals were detected with Illumina BeadArray Reader. Scanned images were converted to signal intensities by BeadStudio expression module (Illumina). Processing of Illumina Microarray, such as background correction, transformation, normalization and identifying differentially expressed genes were performed by lumi package (Du et al., 2008). Differentially expressed probes were filtered (absolute fold change  $\geq 1.5$ ,  $p < 0.05$ ) and multiple probes for a gene were then collapsed into mean value. GSEA (Broad Institute, Cambridge, MA, USA) were performed for comparing the expression of defined gene sets.

## RESULTS

### *Genome-wide identification of in vivo Twist2-binding sites by ChIP-seq*

To identify target genes of Twist2, ChIP-seq was performed with thymocytes isolated from the Twist2-transgenic mice (Oh, 2016). The reason why I took advantage of the Twist2-transgenic mouse system is due to the small number of positively selected DP thymocytes. Since only a small fraction of DP thymocytes (~5%) is positively selected, the amount of primary material available was limited and the use of ChIP-seq to analyze Twist2 occupancy in post-DP thymocytes was negated. In the Twist2-transgenic mice, ectopic expression of Twist2 is under the control of proximal Lck promoter, thereby allowing a majority of DP thymocytes to express Twist2 in the absence of a TCR signal (Oh, 2016). Considering Twist2 expression is induced in DP thymocytes immediately after receiving TCR-mediated positive-selection signal, Twist2-transgenic thymocytes could be assumed to mimic post-DP thymocytes, while providing experimentally favorable condition. Formaldehyde-fixed chromatin of the Twist2 transgenic whole thymocytes was immunoprecipitated with Twist2 antibody, and the recovered DNA was analyzed using SOLiD 3.5 next-generation sequencing platform. Approximately 3.8 and 4.2 million uniquely mappable sequence reads were generated from the Twist2-enriched ChIP sample and the accompanying control (Input DNA) sample, respectively (Table 1).

**Table 1. Properties of sequenced libraries**

Whole thymocytes of Twist2-transgenic mice were chromatin immunoprecipitated with Twist2 antibody, and the recovered DNA was analyzed with SOLiD 3.5 next-generation sequencing platform. The number of sequence reads from the Twist2-enriched ChIP sample and the Input DNA was summarized. The read mapping rates were indicated in percentages.

Library	Read type	Read length	Total number of reads	Reads aligned to mm9
Twist2 ChIP	Single end	50 bp	35,931,122	18,380,813 (51.16%)
Input	Single end	50 bp	39,966,916	22,431,982 (56.13%)

These sequence reads were aligned to the mouse reference genome (mm9), and putative Twist2-binding sites were determined using the Model-based Analysis for ChIP-seq (MACS) (Zhang et al., 2008). As a result, a total of 7,260 significant Twist2 ChIP-seq peaks were identified (Table 2). Twist2 is known to bind to the proximal enhancer region of *Cd4* loci (Lee, 2005). In order to evaluate the Twist2 ChIP-seq results, sequenced reads were overlaid as a track in the IGV (Robinson et al., 2011) and the presence of Twist2 ChIP-seq peaks was checked at the genomic loci of *Cd4*. As expected, Twist2 binding at the proximal enhancer region of *Cd4* was confirmed (Figure 8).

Genomic binding by transcription factor is targeted to preexisting foci of accessible chromatin. Genome-wide open regions in chromatin structure can be probed at a high resolution by partial DNaseI digestion of chromatin combined with massive parallel sequencing (DNase-seq) (Boyle et al., 2008). Recently, Vanhille et al. identified a set of DNaseI hypersensitive sites in the mouse DP thymocytes genome via conducting DNase-seq with purified DP thymocytes (Vanhille et al., 2015). These datasets were employed to sort out Twist2 ChIP-seq peaks that were located at the open chromatin regions in the DP-thymocytes genome. As a result, a total of 3,822 Twist2 peaks that intersected with DNaseI hypersensitive sites are identified and considered as highly confident Twist2-binding sites (Table 2).

Next, genome-wide distribution of Twist2-binding sites relative to the position of known and predicted genes was analyzed using GREAT tools (McLean et al., 2010).



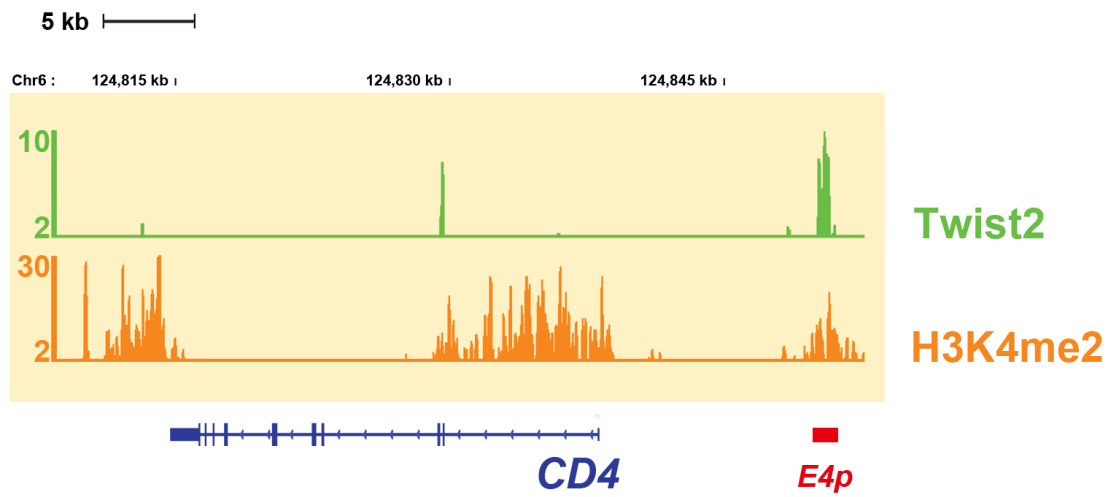
**Table 2. Properties of identified peaks (1)**

The uniquely mappable sequence reads were aligned to the mouse reference genome (mm9), and putative Twist2-binding sites were determined using Model-based Analysis for ChIP-seq (MACS). The data from the DNase-seq analysis of purified DP thymocytes conducted by Vanhille et al. were analyzed. Twist2 ChIP-seq peaks that intersected with the DNaseI hypersensitive sites are considered as highly confident Twist2-binding sites.

Factor	MACS peaks	Intersection with DNase-seq peak regions	Thymocytes population	Ref.
<b>DNase1</b>	56,395		DP	Vanhille et al., 2015
<b>Twist2</b>	7,260	3,822	whole	

### **Figure 8. Evaluation of the Twist2 ChIP-seq results**

A representative view of the Twist2 ChIP-seq analysis is depicted. The Twist2 ChIP-seq peaks were mapped to the July 2007 Mouse Genome Assembly (mm9). The Twist2 and H3K4me2 ChIP-sequenced reads were overlaid as tracks in the Integrative Genomics Viewer (IGV). *In vivo* Twist2 binding at the CD4 proximal enhancer (E4p) region was identified. The area encompassing *Cd4* locus is displayed.



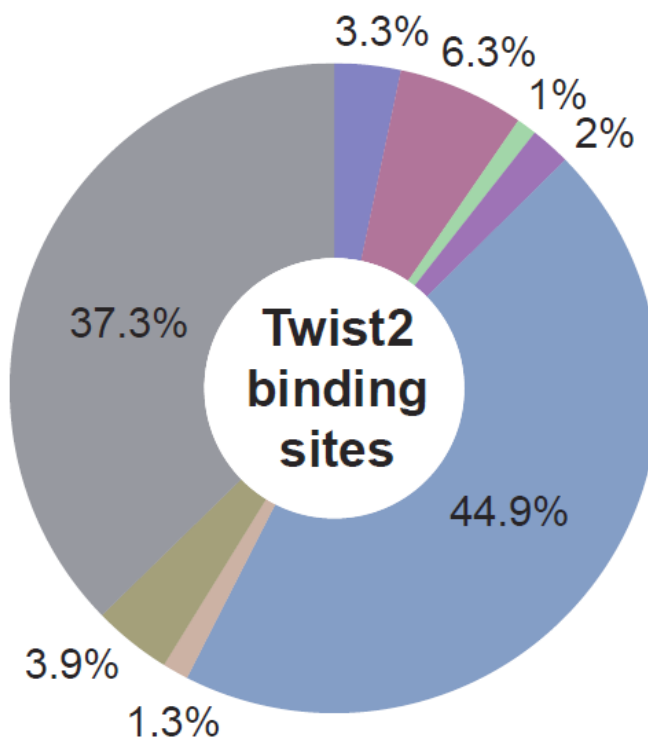
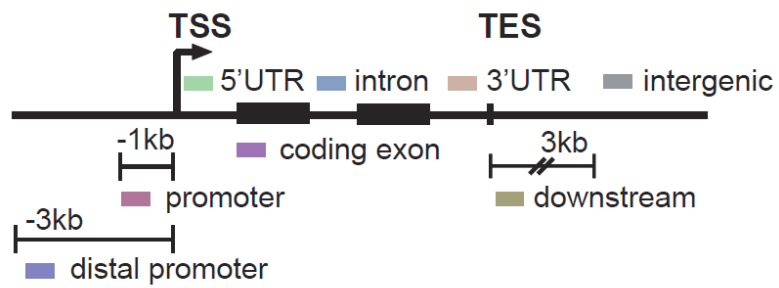
As demonstrated in Figure 9, the vast majority of Twist2-binding sites were positioned in ‘distal intergenic’ and ‘intronic’ regions. This preference of binding is not unusual, since the annotated ‘distal intergenic’ and ‘intronic’ regions are quite abundant in the mouse genome (52.1% and 39.3%, respectively). In particular, Twist2-binding sites were substantially enriched within ‘promoter’, ‘5’ untranslated region (UTR)’, and ‘distal promoter’ when the relative portion of each genomic region were considered (Figure 10). The gene-proximal location of Twist2-binding sites was also revealed by plotting distance of each peak relative to a transcription start site (TSS) for a known gene (Figure 11).

To determine evolutionary conservation of Twist2-binding sites, 3 kb genomic sequences centered at the peak summit of Twist2-binding sites were analyzed with the average phylogenetic conservation score (PhastCons score) across 28 vertebrate species (Siepel et al. 2005). As shown in Figure 12, a prominent peak of conservation was identified in the center of Twist2-binding sites. The presence of strong evolutionary conservation over Twist2-binding sites suggests that these regions could be under selective pressure and therefore might be functionally important (Visel et al. 2007).

Next, motif analysis was performed. Three motifs were significantly enriched ( $E\text{-value} < 1.0 \times 10^{-30}$ ) in Twist2-binding sites. The canonical E-box sequence (5'-CABMTGB-3') was ranked as the top-scoring motif (Figure 13). It was detected in 91.63%(3,502/3,822) of Twist2-binding sites. The additional over-represented motifs are 5'-MGGAAR-3' (37.02%, 1,415/3,822) and 5'-TGTGGTT-3' (10.99%, 420/3,822),

**Figure 9. Genome-wide distribution of Twist2-binding sites relative to known and predicted genes in the mouse genome**

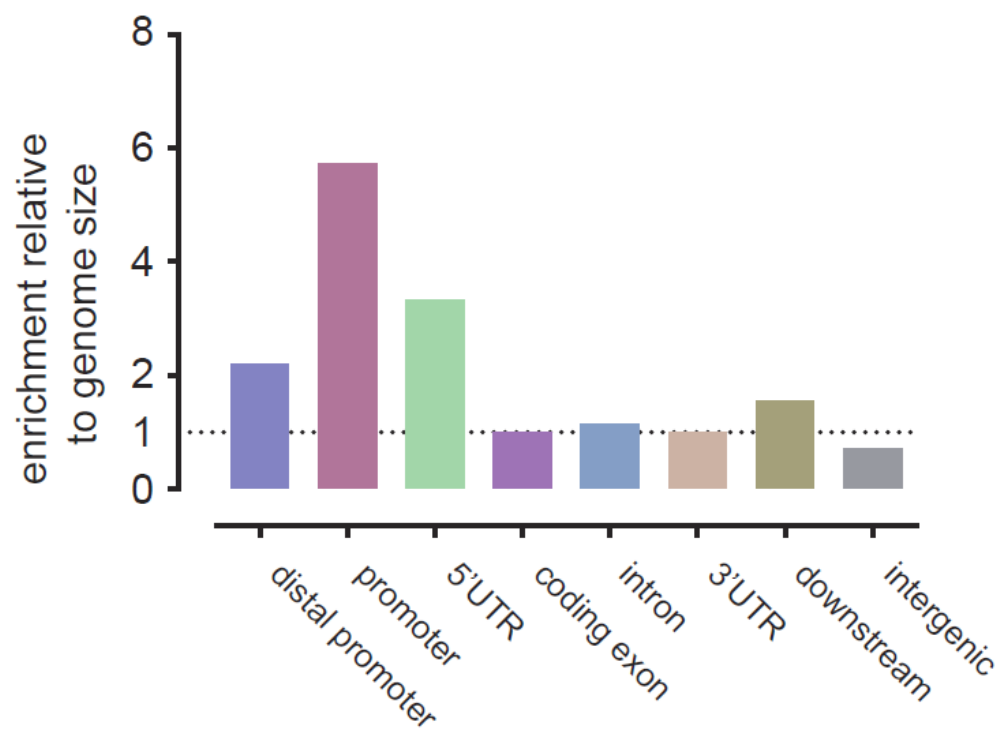
Pie chart represents the genomic distribution of Twist2-binding sites. Defined regions in the mouse genome are depicted as a color code. ‘Promoter’ are defined as regions from -1kb relative to the transcription start site (TSS); ‘distal promoter’ are the regions encompassing -3kb and -1kb with respect to the TSS; Gene bodies are further categorized into untranslated regions (3’ and 5’UTRs), ‘coding exon’ and ‘intron’; ‘downstream’ are defined as regions +3kb from the transcription end site (TES); All the remaining regions are classified as ‘intergenic’. The percentages of given regions in total Twist2-binding sites are indicated.



**Figure 10. Enrichment of Twist2-binding sites in the defined genomic regions**

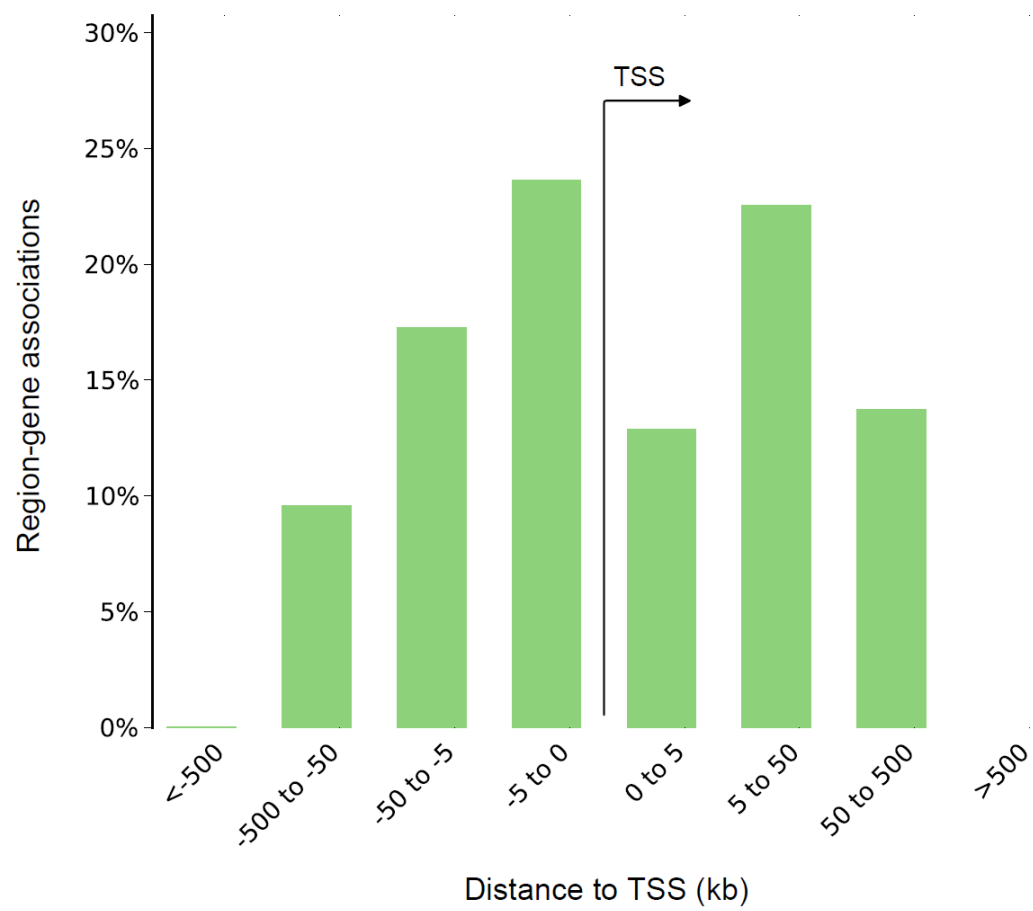
Each bar represents the enrichment of Twist2-binding sites relative to the proportion of corresponding regions in the genome. Twist2-binding sites were substantially enriched within ‘promoter’, ‘5’UTR’, and ‘distal promoter’ regions.





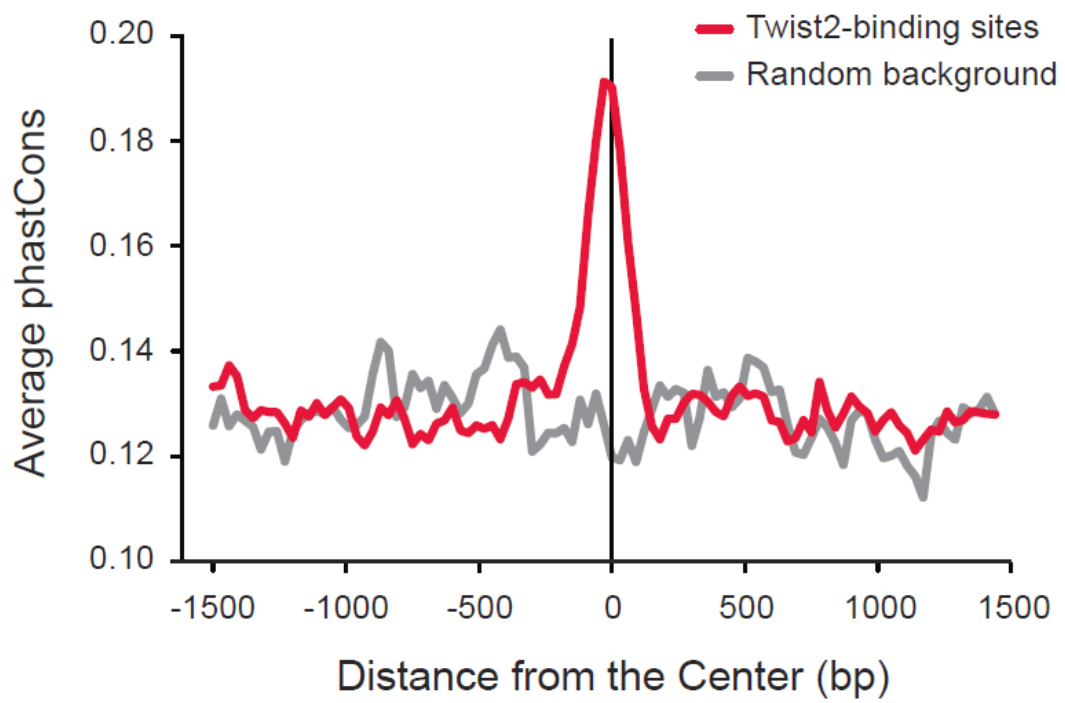
**Figure 11. Distance of Twist2-binding sites relative to the TSS of annotated genes**

Distance of each peak relative to the TSS for a known gene is plotted. The gene-proximal location of Twist2-binding sites was revealed.



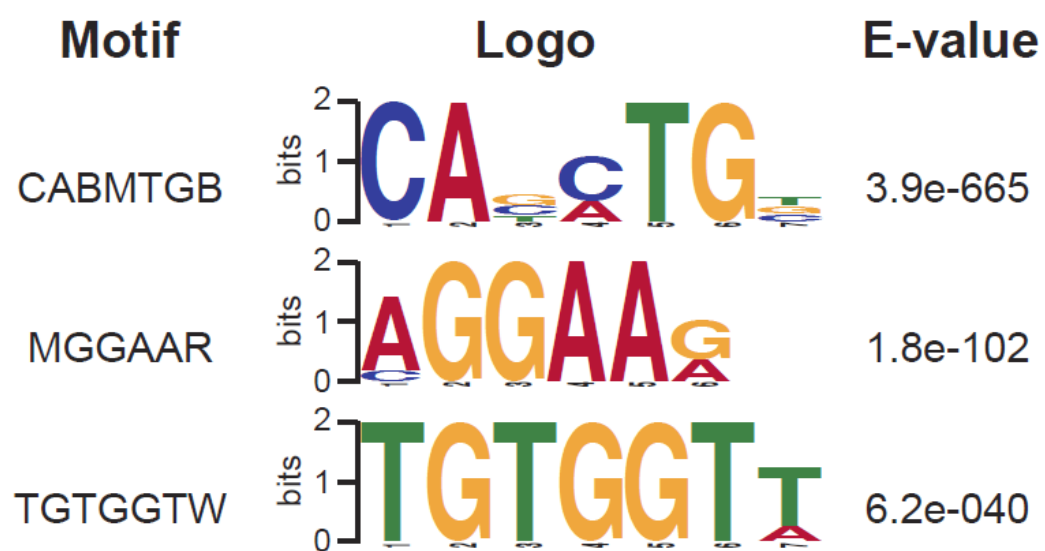
**Figure 12. Evolutionary conservation of Twist2-binding sites**

Average PhastCons vertebrate-conservation score at each base centered at the peak summit of Twist2-binding sites is plotted. The conservation-score plot of random genomic sequences is included as a negative control.



**Figure 13. Consensus DNA sequences associated with Twist2 occupancy**

Twist2 consensus binding sequences predicted by the motif-discovery algorithm are presented as sequence LOGOs. DNA sequences located 100 bp upstream and downstream from the center of Twist2-binding sites were analyzed by DREME 4.3. Letter size indicates nucleotide frequency. Smaller E-values indicate greater enrichment for a given motif than for randomly selected motifs.



which correspond to the consensus DNA sequence for the binding of ETS and Runx family of transcription factors, respectively. To determine the relationship between Twist2 occupancy and its associated *cis*-regulatory elements, the relative distribution of three motifs around the center of Twist2-binding site were analyzed. 5'-CABMTGB-3' motif was enriched in the vicinity of the center of Twist2-binding sites, but the others were distributed evenly (Figure 14).

E box is the consensus DNA sequence to which bHLH transcription factors bind as homo or heterodimers. In DP thymocytes, E proteins (HEB and E2A, members of bHLH transcription factors) are highly expressed and regulate expression of critical factors by binding to the E-box sequences located at their regulatory regions. Since Twist2 dimerize preferentially with E proteins to form heterodimers, I hypothesized that Twist2 may be targeted to the E-box sites pre-occupied by E proteins after being induced in post-selection DP thymocytes.

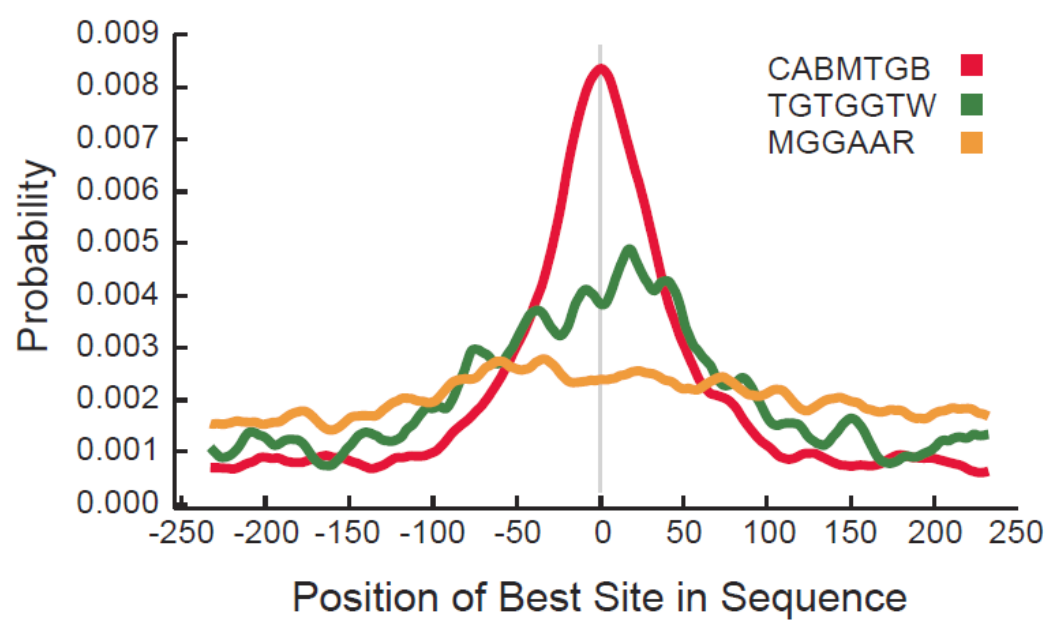
#### *Twist2 and HEB bind to overlapping sites on the DP-thymocytes genome*

To validate this tentative hypothesis, a recent study performed by Vanhille et al., which identified HEB-binding sites in DP thymocytes was exploited (Vanhille et al., 2015). The HEB ChIP-seq data were interrogated and 1,655 HEB-binding sites were identified (Table 3). These sites were then compared with Twist2-binding sites.



**Figure 14. Relative distribution of motifs across Twist2-binding sites**

Frequencies of consensus DNA sequence motifs are plotted as a function of distance from the center of Twist2-binding sites. 500 bp of genomic sequences centered at the peak summit of Twist2-binding sites were analyzed. The canonical E-box sequence (5'-CABMTGB-3') is located in close proximity to the center of Twist2-binding sites. The other overrepresented motifs, 5'-MGGAAR-3' and 5'-TGTGGTT-3', were revealed to distribute evenly across Twist2-binding sites.



**Table 3. Properties of identified peaks (2)**

The uniquely mappable sequence reads were aligned to the mouse reference genome (mm9), and the putative HEB-binding sites were determined using Model-based Analysis for ChIP-Seq (MACS). HEB ChIP-seq peaks that intersected with DNaseI hypersensitive sites were considered as highly confident HEB-binding sites.

Factor	MACS peaks	Intersection with DNase-seq peak regions	Thymocytes population	Ref.
<b>DNase1</b>	56,395		DP	Vanhille et al., 2015
<b>HEB</b>	15,895	1,655		

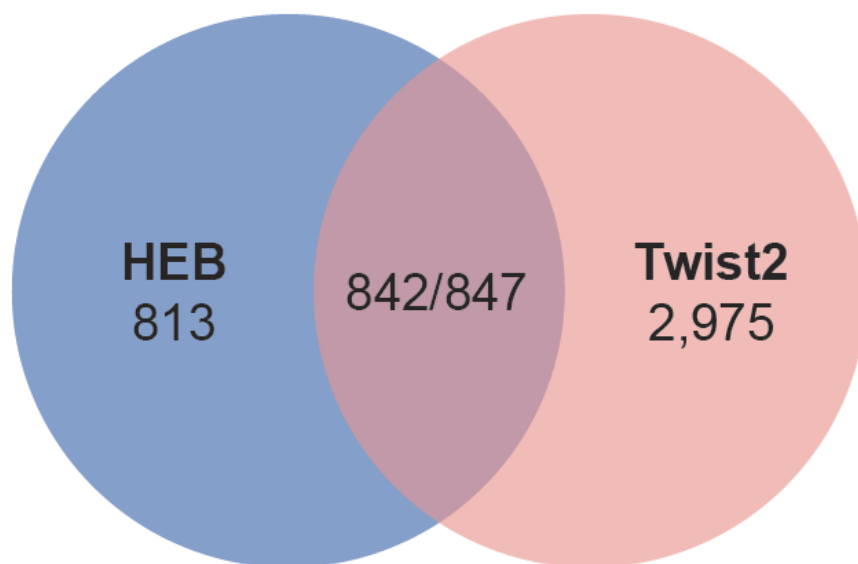
As a first step, the number of regions shared by HEB and Twist2 were analyzed. Strikingly, the binding profiles of HEB and Twist2 strongly overlapped. Of HEB-binding sites, 51% (842/1,655) were overlapped with Twist2-binding sites (Figure 15). In contrast, 22% (847/3,822) of Twist2-binding sites were overlapped with HEB-binding sites. Next, the extent of overlapping between two factors was investigated. As shown in Figure 16, 92.76% of HEB-binding sites were intersected with Twist2-binding sites more than half of their length. In case of Twist2, 86.42% of Twist2-binding sites were intersected with HEB-binding sites more than half of their length. In conclusion, Twist2 binds to about 50% of genomic loci pre-occupied by HEB during thymocyte selection with substantial overlapping.

#### *Twist2 negatively regulates HEB target genes during thymocyte selection*

To identify putative genes directly regulated by Twist2 and/or E proteins in DP thymocytes, the occupancy peaks were linked to their adjacent genes. At first, Twist2- and HEB-binding sites were categorized into the three groups: those binding of both Twist2 and HEB (composite sites), those binding of Twist2 or HEB alone. And then, these binding sites were assigned to particular genes if the site was present within 15kb upstream and 5kb downstream of the annotated transcription start sites. As results, 239 genes were associated with the composite sites, and 812 and 191 genes were associated

**Figure 15. Common chromatin occupancy by Twist2 and HEB**

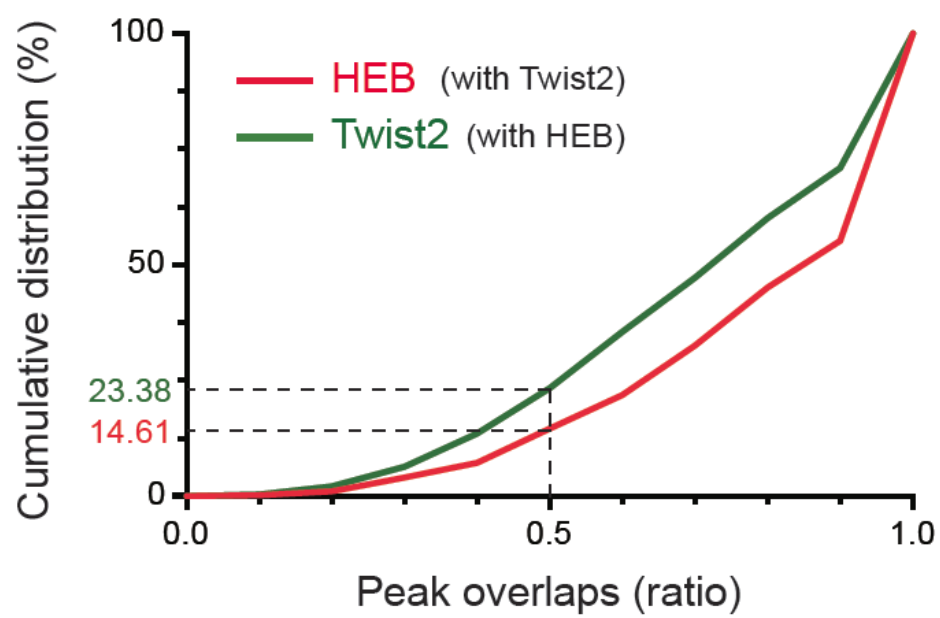
The numbers of genomic loci occupied by Twist2 and/or HEB are depicted as a Venn diagram. Of the 1,655 HEB-binding sites, 51% (842) were overlapped with Twist2-binding sites. In contrast, 22% (847) out of the 3,822 Twist2-binding sites were overlapped with HEB-binding sites.



**Figure 16. Extent of overlapping between Twist2- and HEB-binding sites**

Cumulative frequencies with increasing proportion of overlaps are plotted: HEB (red curve) and Twist2 (green curve). 92.76% of HEB-binding sites and 86.42% of Twist2-binding sites were intersected with each other more than half of their length.





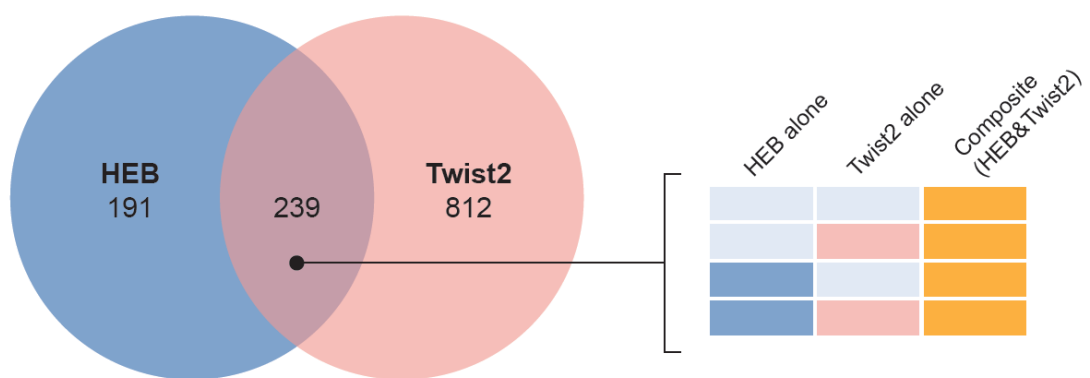
with the ‘Twist2 alone’ and the ‘HEB alone’ sites, respectively (Figure 17). As expected, 55.6% (239 out of 430) of the HEB target genes had composite sites in their regulatory region, at which Twist2 also bound.

In the aforementioned peak analysis, it was suggested that Twist2 might regulate E protein target genes during thymocyte selection. To evaluate how the Twist2 occupancy correlates with the differential gene expression in pre-selection DP (hereafter, pre-DP, where Twist2 level is low) and post-selection DP (hereafter, post-DP, where Twist2 level is elevated) thymocytes, the global gene-expression profiles before and immediately after thymocyte selection was compared using microarray. Two surface markers, CD5 and TCR $\beta$ , were used for gating pre-DP (CD5<sup>med</sup>TCR $\beta$ <sup>med</sup> CD4+CD8+) and post-DP (CD5<sup>high</sup>TCR $\beta$ <sup>med</sup> CD4+CD8<sup>lo</sup>) thymocytes (Figure 18A). Before microarray analysis, the induction of Twist2 expression in post-DP thymocytes was confirmed using these two FACS-purified subsets (Oh, 2010). As expected, *Twist2* mRNA level was elevated in post-DP thymocytes, suggesting that Twist2 plays roles in post-DP thymocytes during thymocyte selection (Figure 18B). Biotinylated RNA prepared from pre-DP and post-DP thymocytes were hybridized to Illumina Mouse Ref-8 Expression BeadChip. Differential expression for each probe was calculated and then filtered with the criteria: fold change  $\geq 1.5$ , and  $p < 0.05$ . Multiple probes for a gene were collapsed into a mean value.

Microarray analysis revealed 2,311 differentially expressed genes (DEG) (Figure 19). Among them, 1,142 genes were down regulated and 1,169 genes were up

**Figure 17. Common target genes of Twist2 and HEB**

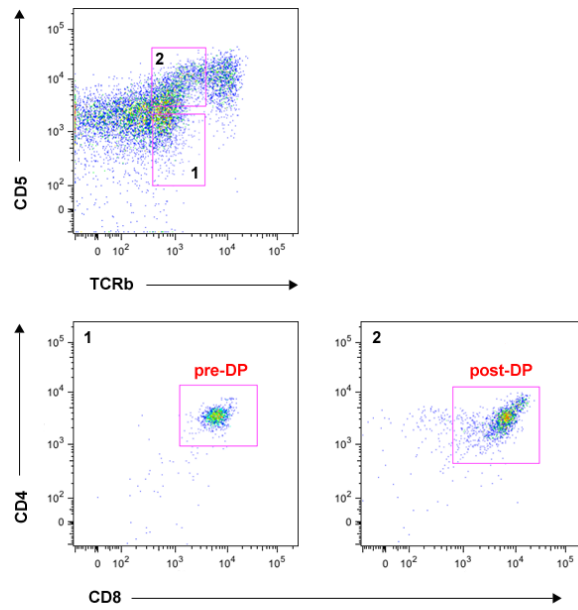
Numbers of genes associated with the composite, the Twist2-alone, and the HEB-alone sites are shown in a Venn diagram.



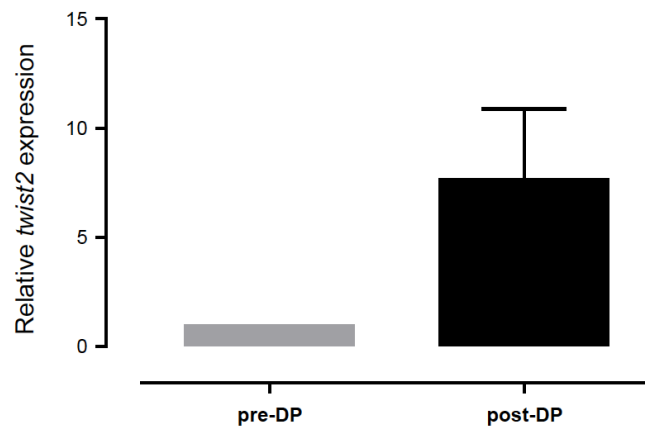
**Figure 18. Purification of cells for mRNA microarray analysis**

(A) Gates show pre-DP ( $CD5^{med}TCR\beta^{med}$ ) and post-DP ( $CD5^{hi}TCR\beta^{med}$ ) subsets used for subsequent analysis. Thymocytes were stained for the expression of CD4, CD8, CD5, and TCR $\beta$  and sorted as fractions pre-DP and post-DP. At first, fractions 1 and 2 were gated on CD5-TCR $\beta$  plot, and then pre-/post-DP were gated as indicated on the CD4-CD8 plot of the fractions 1 and 2. (B) Twist2 is highly expressed in post-DP thymocytes. *Twist2* mRNA level in pre-DP and post-DP thymocytes were quantified by qRT-PCR and compared. *Gapdh* was used for normalization. These are results from three independent experiments.

(A)

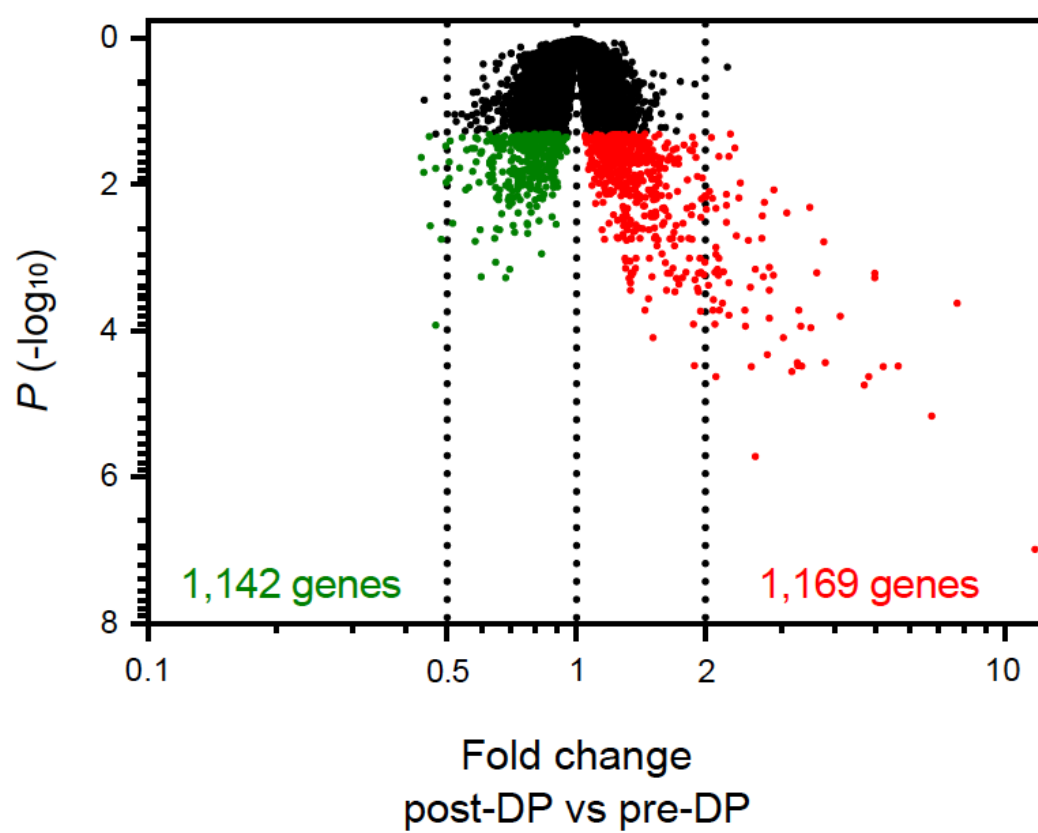


(B)



**Figure 19. Differential gene expression between pre- and post-DP thymocytes**

Microarray analysis comparing gene expression in pre-DP and post-DP thymocytes is presented as a Volcano plot. Vertical axis represents the t-test P-values. Changes in gene expression (horizontal axis) are shown as a ratio of post-DP to pre-DP thymocytes. Green and red dots highlight genes with greater than 1.5-fold change and t test P-value  $< 0.05$ . Black dots show genes remaining after quality filtering.





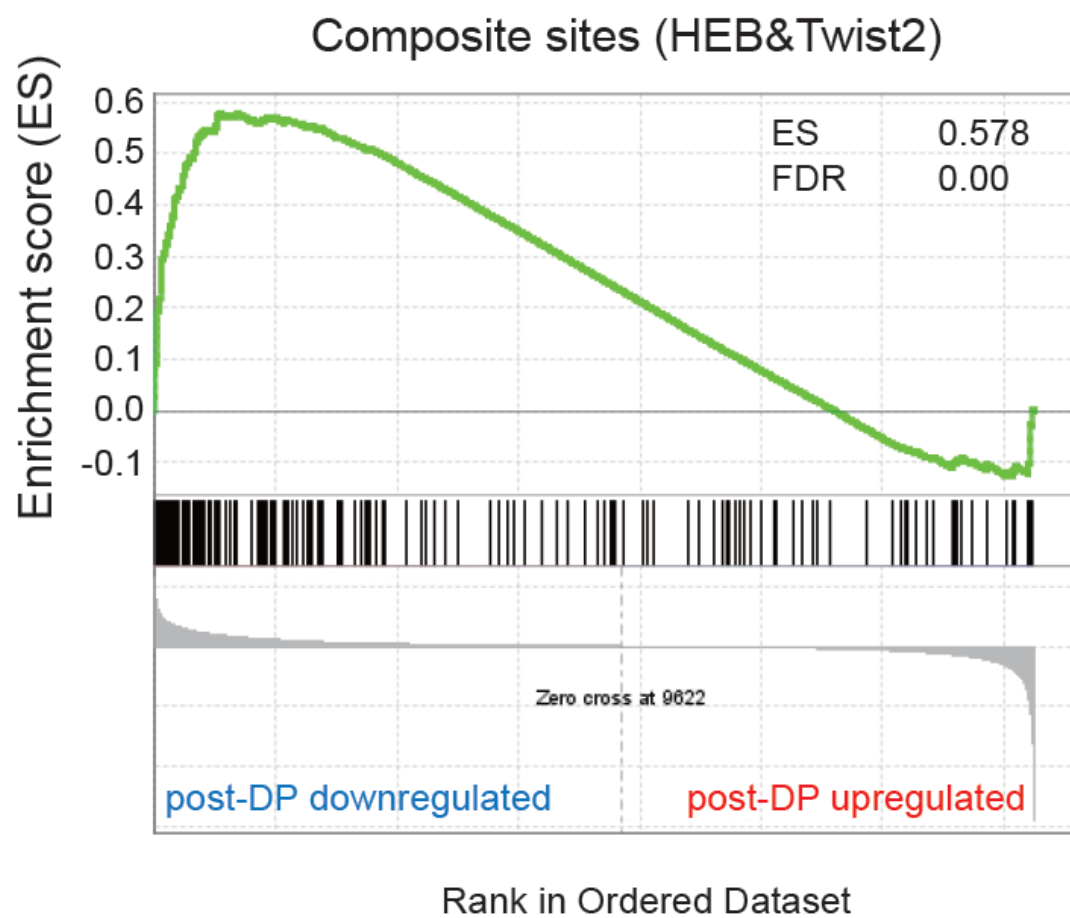
regulated. Many of genes on the DEG list had previously been reported to change during thymocyte selection, including *Rag1/2*, *Bcl2l1*, *Rorc*, *Gata3*, *Egr1*, and *Tox*, thus confirming the general validity of the analysis.

To gain insight on the functional consequences of Twist2 binding to the composite sites during thymocyte selection, Gene-Set-Enrichment Analysis (GSEA) was performed using gene set associated with the composite sites. In the GSEA, all of genes analyzed by microarray were distributed on the x-axis as their rank order of differential expression in post-DP versus pre-DP thymocytes and occurrence of a gene from the given gene set is then scanned for going from the most down regulated to the most up regulated genes. Presence or absence of a gene in the gene set is then scored and plotted on the y-axis with a running enrichment score. This analysis revealed that genes associated with the composite sites were highly enriched among the group of genes down regulated upon thymocyte selection (ES 0.578, FDR 0.00) (Figure 20). On the other hand, there was no tendency in gene expression changes for genes associated with the Twist2 alone sites.

To confirm the expression of genes associated with the composite sites is dependent on E-protein activity, the gene set associated with the composite sites was analyzed by GSEA for overlapping with genes that were preferentially down regulated upon removal of E proteins. For this purpose, I took advantage of microarray data from a recent study, which analyzed the differential gene expression in the *Tcf12<sup>f/f</sup>Tcfe2a<sup>f/f</sup>* CD4Cre<sup>+</sup> versus the *Tcf12<sup>f/f</sup>Tcfe2a<sup>f/f</sup>* CD4Cre<sup>-</sup> DP thymocytes (Jones and Zhuang, 2007).

**Figure 20. Correlation between Twist2 occupancy and the gene expression changes upon thymocyte selection**

GSEA plot indicates the degree to which genes associated with the composite sites are overrepresented at the extreme left (down regulated during thymocyte selection) or right (up regulated during selection) of the entire ranked list. The x-axis represents the rank order of genes according to their differential expression in post-DP versus pre-DP thymocytes. The tick marks below the plot indicate the presence of a gene in the given gene set. Solid bars represent target genes.

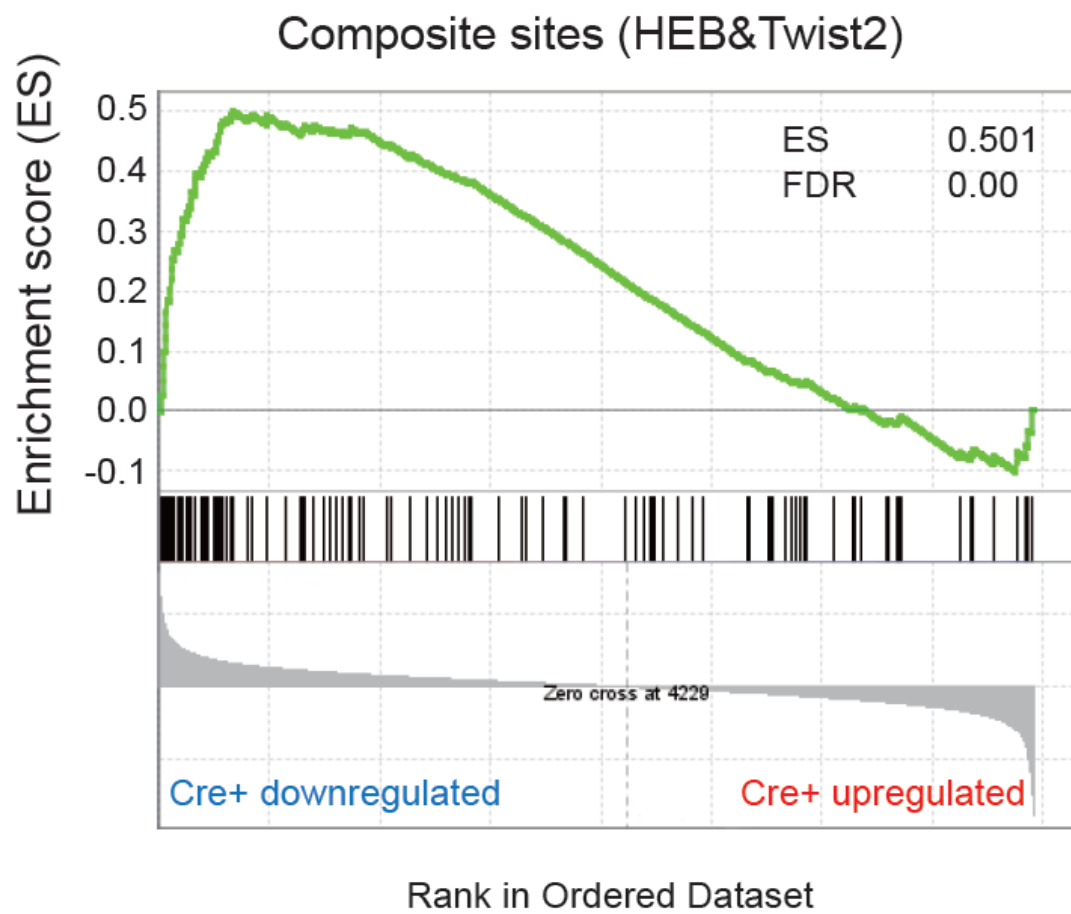


As expected, genes associated with the composite sites were highly enriched among the group of genes down regulated in the  $Tcf12^{f/f}Tcfe2a^{f/f}CD4Cre+$  DP thymocytes (ES 0.501, FDR 0.00) (Figure 21). This indicated that E-protein activity is indeed crucial for the proper expression of genes associated with the composite sites in pre-DP thymocytes. These analyses strongly suggested that Twist2 might negatively regulate E protein target genes during thymocyte selection by binding to the genomic sites preoccupied by E proteins.

Out of the 239 genes associated with the composite sites, 73 genes showed significant differential expression ( $p\text{-value}<0.05$ ,  $FC>1.5$ ) during thymocyte selection. Among them, 23 genes were also differentially expressed ( $FC>1.5$ ) upon deletion of E proteins. When the expression of these 23 genes was compared, a similar trend between the two microarray data was emerged (Figure 22). Genes that were strongly repressed in response to thymocyte selection were also down regulated upon E protein removal (ClusterA), whereas genes highly induced in post-DP thymocytes were also up regulated in the  $Tcf12^{f/f}Tcfe2a^{f/f}CD4Cre+$  DP thymocytes (ClusterB). However, several genes showed inverse profiles in expression between two microarray data (ClusterC and ClusterD). If the co-occurrence of Twist2 binding in HEB-binding sites was functionally important, it should be confirmed that Twist2 is responsible for negatively regulating expression of genes associated with the composite sites during thymocyte selection. For further validation, 18 genes were chosen from ClusterA and ClusterB.

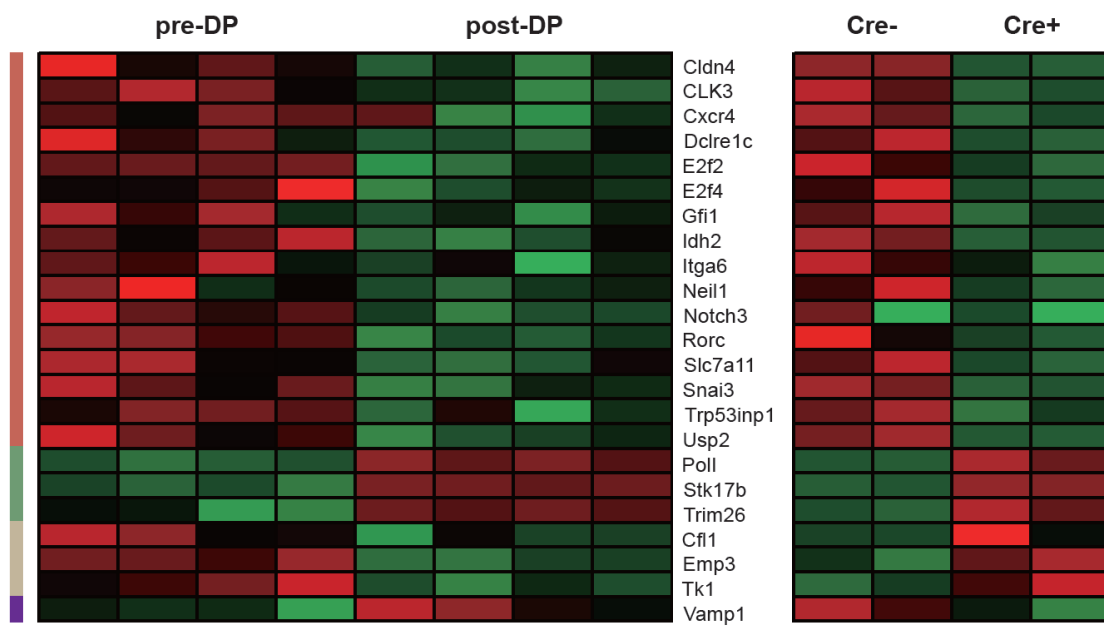
**Figure 21. Correlation between Twist2 occupancy and the gene expression changes upon E protein removal**

GSEA plot indicates the degree to which genes associated with the composite sites are overrepresented at the extreme left (down regulated in  $Tcf12^{f/f}Tcf2a^{f/f}CD4Cre^{+}$  DP thymocytes) or right (up regulated in  $Tcf12^{f/f}Tcf2a^{f/f}CD4Cre^{+}$  DP thymocytes) of the entire ranked list. The x-axis represents the rank order of genes according to their differential expression in  $CD4Cre^{+}$  versus  $CD4Cre^{-}Tcf12^{f/f}Tcf2a^{f/f}$  DP thymocytes. The tick marks below the plot indicate the presence of a gene in the given gene set. Solid bars represent target genes.



**Figure 22. Significant changes in the expression of genes associated with the composite sites during thymocyte selection as well as upon E protein removal**

Expression level of 23 genes associated with the composite sites across samples are normalized per gene and shown as heat maps. Four replicates are shown for pre-DP and post-DP thymocytes. Two replicates are shown for  $Tcf12^{f/f}Tcfe2a^{f/f}CD4Cre^-$  DP and  $Tcf12^{f/f}Tcfe2a^{f/f}CD4Cre^+$  DP thymocytes. Red color represents up regulation and green color represents down regulation in gene expression.





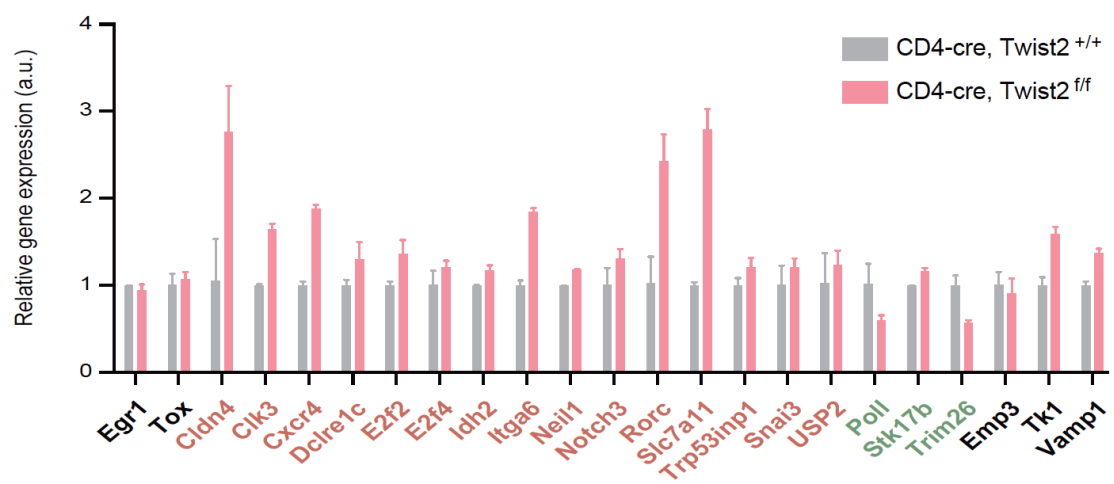
By using DO11.10 TCR-transgenic mouse system, effects of Twist2 knock out on the expression of these genes were assessed (Figure 23). In DO11.10 TCR-transgenic mice, the majority of DP thymocytes have received positive-selection signal, as evidenced by CD69 up-regulation. And the DO11.10 thymus contains a large population of CD4<sup>+</sup>CD8<sup>low</sup> cells that are “intermediate” between immature DP and mature SP (Lundberg et al., 1995). In the absence of Twist2, genes that are known to be down regulated in response to thymocyte selection were up regulated. For example, *Rorc* and *Cldn4*, whose expression dropped dramatically upon receipt of positive-selection signal, were both up regulated in Twist2<sup>f/f</sup>CD4Cre+ DO11.10 cells. In contrast, *Poll* and *Trim26*, which were highly expressed in positively selected DP cells, were down regulated in Twist2<sup>f/f</sup>CD4Cre+ DO11.10 cells. Therefore, Twist2 was proven to be responsible for negatively regulating expression of genes associated with the composite sites during thymocyte selection.

#### *Negative regulation of Rorc expression by Twist2 during thymocyte selection*

To evaluate whether Twist2 directly regulates the expression of E protein target genes by binding to the composite sites, retinoic acid receptor-related orphan receptor gamma t (RORγt) was selected as a representative. RORγt, an isoform of RORγ, is a key player in the establishment of the DP thymocyte phenotype (He, 2002; Jetten et al., 2001;

**Figure 23. Effects of Twist2 knock out on the expression of genes associated with the composite sites**

Expression levels of 18 Twist2 target genes in DO11.10 thymocytes were analyzed by qRT-PCR. DO11.10 TCR<sup>low</sup> population was isolated from the Twist2<sup>f/f</sup>CD4Cre+ DO11.10 and the Twist2<sup>+/+</sup>CD4Cre+ DO11.10 thymuses. Data were normalized to  $\beta$ -actin expression. The expression of each gene in the Twist2<sup>+/+</sup>CD4Cre+ DO11.10 mice was set to 1.

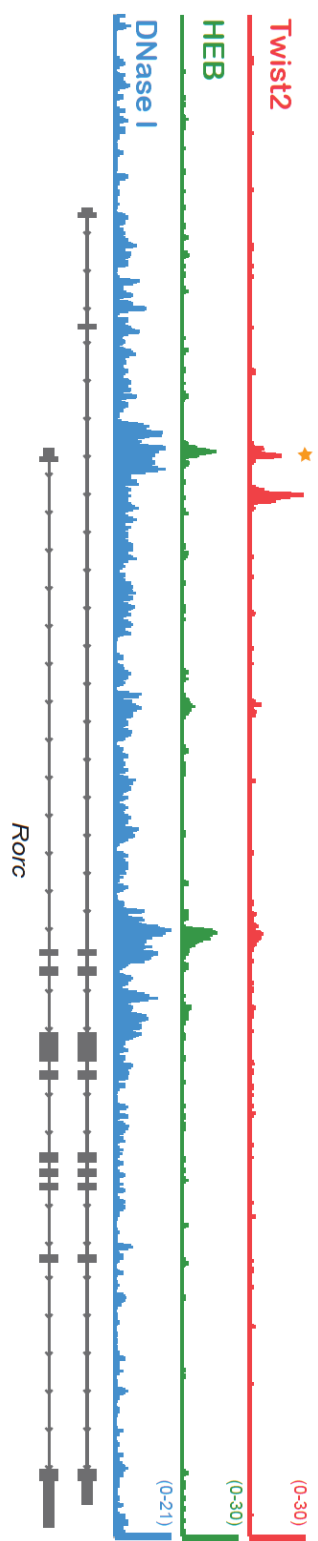


Jetten and Ueda, 2002). ROR $\gamma$ t is transcribed from *Rorc* loci and its expression is confined to DP thymocytes and lymphoid tissue inducer cells (Eberl et al., 2004; He et al., 2000). ROR $\gamma$ t-deficient mice lack lymph nodes and also have DP thymocytes that show signs of increased cell-cycle activity (Sun et al., 2000). DP thymocytes in ROR $\gamma$ t-deficient mice also have poor survival, resulting in an abnormally small thymus, despite the enhanced proliferation (Kurebayashi et al., 2000; Sun et al., 2000). These results suggest that ROR $\gamma$ t inhibits proliferation and promotes survival of DP thymocytes until TCR-mediated positive selection is transmitted.

*Rorc* encodes both ROR $\gamma$  and ROR $\gamma$ t. These gene products share exons 3–11, but ROR $\gamma$ t does not use exons 1 and 2 of *Rorc*, replacing them with the unique 1 $\gamma$ t exon (Figure 24; He et al., 1998; Medvedev et al., 1997; Villey et al., 1999). The major ROR $\gamma$ t TSS is located just upstream of the 1 $\gamma$ t exon and base -2,242 to +109 relative to the ROR $\gamma$ t TSS bears substantial promoter activity (Xi et al., 2006). Xi et al. intensively investigated ROR $\gamma$ t promoter and found two essential E-box sites (Xi et al., 2006). Binding of E proteins to these E-box sites enhances ROR $\gamma$ t expression. When HEB and Twist2 ChIP-sequenced reads were overlaid onto *Rorc* genomic loci, HEB and Twist2 were found to occupy in close proximity to the ROR $\gamma$ t TSS (Figure 24). Interestingly, the composite binding site within *Rorc* promoter (base -176 to +78 relative to the ROR $\gamma$ t TSS) contains two E-box sites that are reported by Xi et al. These E-box sites are highly conserved across placental mammals (Figure 25).

**Figure 24. Composite binding of Twist2 and HEB to the ROR $\gamma$ t TSS**

The Twist2 and HEB ChIP-seq peaks were mapped to the July 2007 Mouse Genome Assembly (mm9) and the region encompassing *Rorc* locus was displayed. In accordance with previous report (Xi et al., 2006), *in vivo* HEB binding near the ROR $\gamma$ t TSS was confirmed. Twist2 binding was identified within the HEB-binding sites.



**Figure 25. Two highly conserved E-box sites within ROR $\gamma$ t promoter region**

Genomic sequences which correspond to the base -176 ~ +78 relative to the mouse ROR $\gamma$ t TSS were retrieved from four placental mammals (Mouse, Cow, Horse and Human) and aligned. Blue box highlights the ROR $\gamma$ t TSS located just upstream of the 1 $\gamma$ t exon. The two E-box sites, which are essential for the ROR $\gamma$ t promoter activity, are highlighted in red boxes (Xi et al., 2006). These E-box sites are revealed to be highly conserved. An E-box site in yellow box has been reported to be dispensable for the ROR $\gamma$ t promoter activity (Xi et al., 2006) and is not evolutionarily conserved.

Horse 407 GTCATCTTACCCGCTGCACCCAGGGGGGTGAGGGGGTGGAGT--GGGGGGTGTGA 464  
 Human 414 GTCATCTACTCACTGCACCC-----TT--GGGGGTGGGT---GGGGGGTGTCA 460  
 Cow 403 GTCATCTACCCGCTGCACCCAGGGGGGT--GGGGGTGGAGTTGGGGGGGTGTCA 461  
 Mouse 353 GTCATCATACCCAATGCACCTC-----T---GGGGT----T---GGGGGGCTGTCA 394

Horse 465 CTCGGCCACCTGTGTGGTGCAGGCTTAAACCCCCCTGCCAGAAGCGTTGGGGGAGAG- 523  
 Human 461 CTTGGCCACCTGTGTGGTGCAGAGCTTAAACCCCCAGTCCAGAAGCACTGGGGGAGAGA 520  
 Cow 462 CTCGGCCACCTGTGTGGTGCAGATCTTAAACCCCCGGCCCAGAAGCATTGGGGGAGAG- 520  
 Mouse 395 CTTGGCCACCTGTGTGGAGCAGAGCTTAAACCCCCCTGCCAGAAACACTGGGGGAGAG- 453

Horse 524 -CTAGGTGCCGAGCTGCAGGCTGAGAC-TCTGCTGAGAGGATCTGGCCCCGACCCC-ACT 580  
 Human 521 GCTAGGTGCAGAGCTTCAGGCTGAGGC-GCTGCTGAGAGGGCCTCGCCCC-GCCTCTGCC 578  
 Cow 521 -CTAGGTGCAGAGCTGCAGGACTGAGGC-TCTGCCAAGAGGGCCTGGCCCCACCCCTGCT 578  
 Mouse 454 -CTTTGTGCAGATCTAAGGGCTGAGGCACCCGCTGAGAGGGCTTCACCCC-ACCTCCACT 511

Horse 581 GCAGCCGCGCCCCATCCCTGGACCACCCCGGCTG---AGGACAGGGAGCCAAGGC-AG 636  
 Human 579 GCAGCTGCACCCCACTCCTGGACCACCCCTGCTGAGAAGGACAGGGAGCCAAGGCCGG 638  
 Cow 579 GTAGCTGCGCCCCATCCCTGGACCACCC--GGCTGAGAAGGACAGGGAGCCAGGC-GG 636  
 Mouse 512 GCAGCTGTGTGCTGT-CCTGGGCTACCCT--ACTGAGGAGGACAGGGAGCCAAGTT--- 565

Horse 637 CAAAGCCGAAGGCTCAGCCATGAGAAGTAAGTGAATGGGGCCACCTGGGAGCGGGAGAAG 696  
 Human 639 CAGAGCC-AAGGCTCAGTCATGAGAAGTAAGTGAATGGGGCCACCTGGGGGCGGG-GGAG 696  
 Cow 637 CAAAGCCAGGACTCAGTCATGAGAAGTAAGTGAATGGG-CCTCCTGGGAGCGGG-GAAG 694  
 Mouse 566 -----CTCAGTCATGAGAAGTAAGTGAATGGGGGCATCCGGTCATGGG-GGAG 612

Horse 697 CCTGGGTCTTGCCATCAT-CCCGGGCCCTCTGGAAAGGAAGGGTGACTTGGGAACAGTGT 755  
 Human 697 CCTGGACCCTGTCGTCAC-----CCCTCTGGAAAGGA-GGATGTCTTGGGAACAGTGT 748  
 Cow 695 CCTGGGGCCCATCATATTCCCGGGCCTTCTGAAAAGGA-----GGGAACAGTGT 744  
 Mouse 613 CCTGGGTCCTGTCAACATTCTAGGCCGCTGAATAGGA-----GTGTATCTT 660

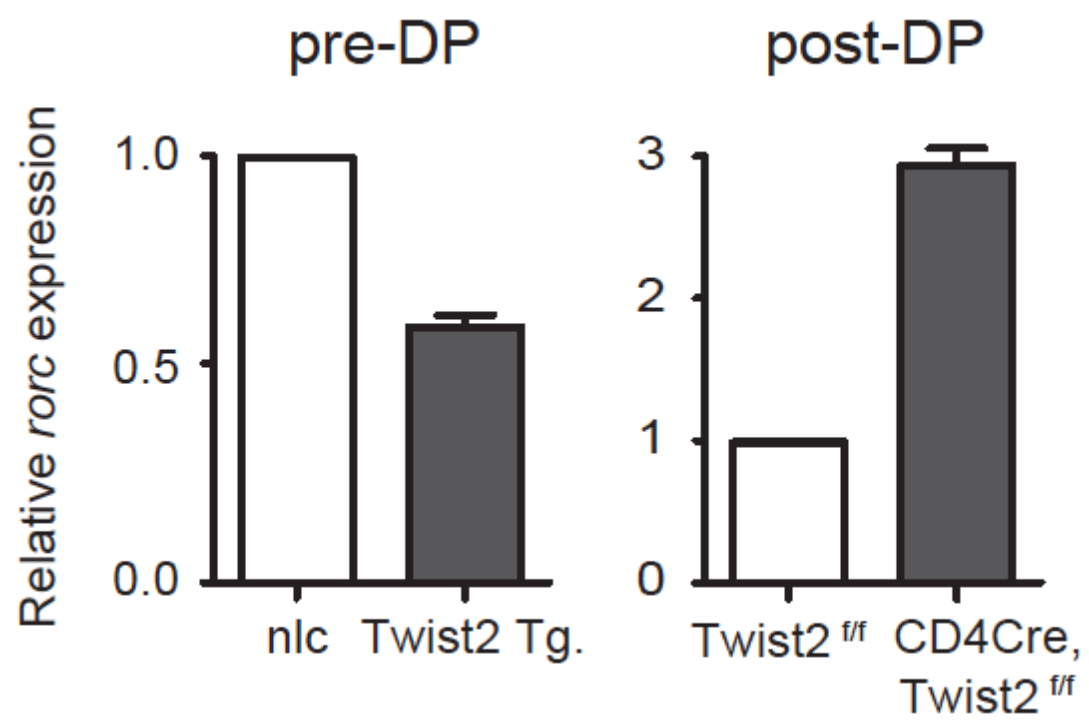


ROR $\gamma$ t is highly expressed in DP thymocytes but is strongly repressed during thymocyte selection. E proteins are responsible for ROR $\gamma$ t expression. I speculated that Twist2 might act as a repressor for ROR $\gamma$ t expression via direct binding to the composite site pre-occupied by E proteins. To confirm this hypothesis, ROR $\gamma$ t expression in DP thymocytes from Twist2-transgenic mice and its littermate control mice were compared. As shown in Figure 26, relative expression of *Rorc* is reduced by nearly 60% in Twist2 transgenic DP thymocytes. In contrast, a 3-fold increase in *Rorc* mRNA is revealed when Twist2 is knocked out in DP thymocytes (Twist2<sup>f/f</sup>CD4Cre+), suggesting role of Twist2 in negative regulation of ROR $\gamma$ t expression in DP thymocytes that has recently experienced positively selecting TCR signal.

Next, the possibility whether Twist2 is inhibitory for ROR $\gamma$ t promoter activity was checked. ROR $\gamma$ t promoter construct (-2,242 to +109 relative to the ROR $\gamma$ t TSS) were co-transfected into 293T cells with increasing amounts of a plasmid encoding Twist2. Twist2 was indeed able to target ROR $\gamma$ t promoter to repress its activity (Figure 27). By performing ChIP-qPCR, binding of Twist2 to ROR $\gamma$ t promoter region was also confirmed (Figure 28). These results strongly supported that Twist2 is responsible for the repression of *Rorc*, a target of E proteins, during thymocyte selection.

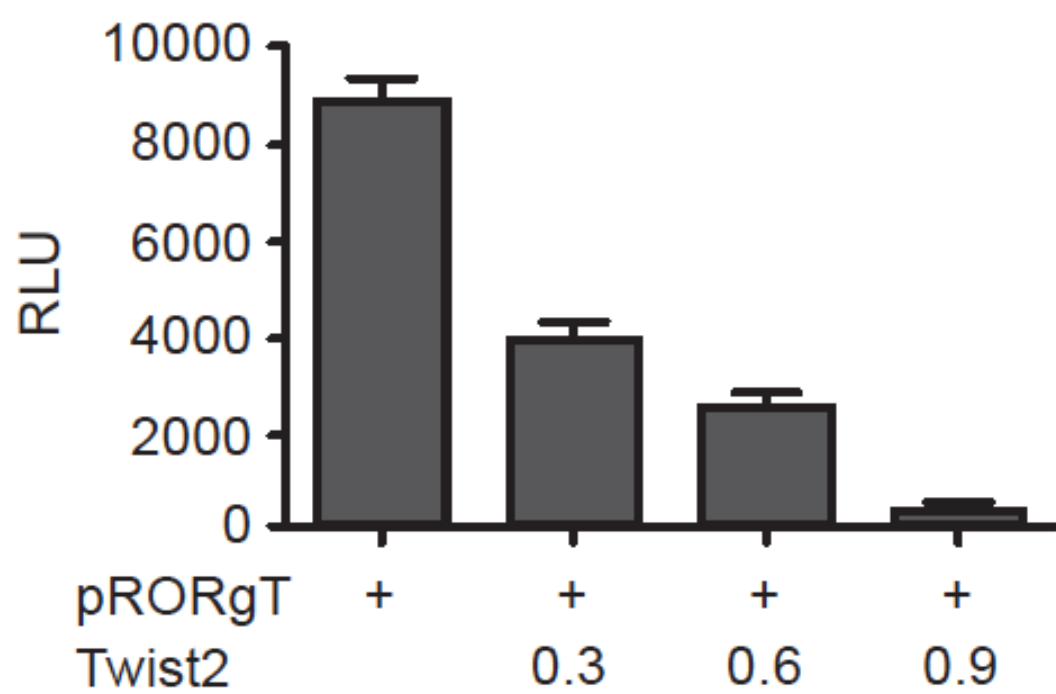
**Figure 26. Negative regulation of ROR $\gamma$ t expression by Twist2**

ROR $\gamma$ t expression in pre-DP thymocytes from Twist2-transgenic mice and its littermate control mice were compared (Left panel). ROR $\gamma$ t expression in post-DP thymocytes from Twist2-KO mice (Twist2<sup>f/f</sup>CD4Cre+) and its control mice (Twist2<sup>f/f</sup>CD4Cre-) were compared (Right panel). *Rorc* mRNA was measured by qRT-PCR and normalized to  $\beta$ -actin expression. Twist2 negatively regulates ROR $\gamma$ t expression in DP thymocytes.



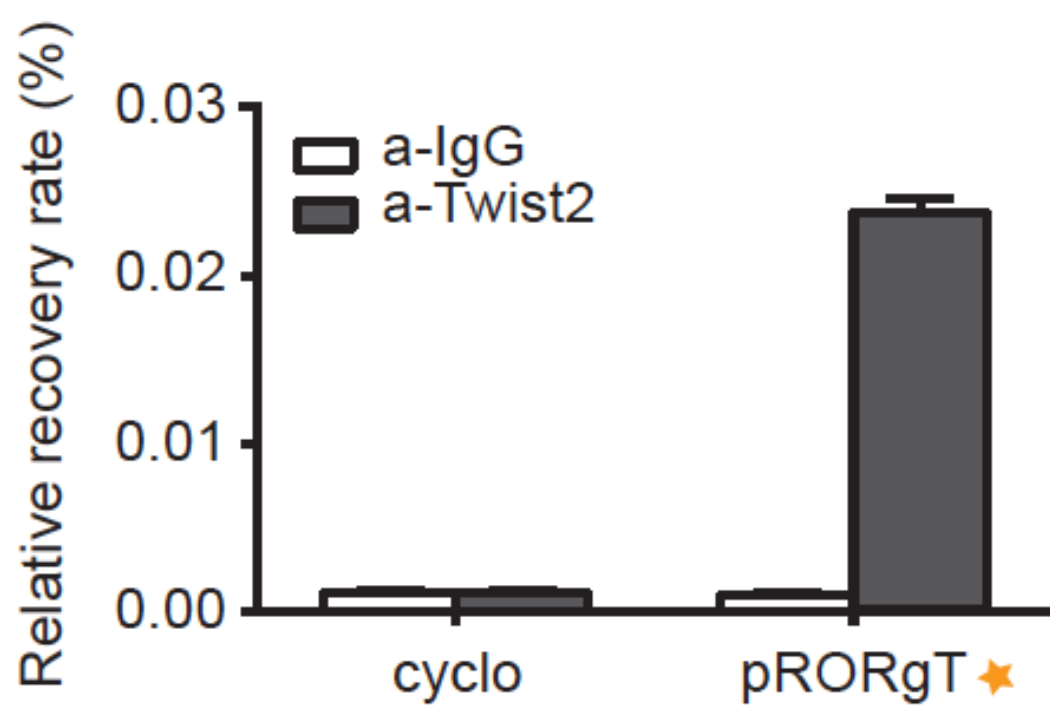
**Figure 27. Negative regulation of ROR $\gamma$ t promoter activity by Twist2**

Overexpression of Twist2 represses ROR $\gamma$ t promoter activity. The luciferase reporter constructs bearing ROR $\gamma$ t promoter (-2,242 to +109 relative to the ROR $\gamma$ t TSS) region was cotransfected into 293T cells with increasing amount (0, 0.3, 0.6, and 0.9  $\mu$ g) of Twist2 expression vector and the luciferase activity was analyzed 48 hr post-transfection. The data show the means  $\pm$  SD of a triplicate experiment and represents three similar experiments.



### **Figure 28. Direct binding of Twist2 to ROR $\gamma$ t promoter**

Twist2 occupancy at ROR $\gamma$ t promoter in Twist2-transgenic thymocytes was evaluated by ChIP analysis. Whole thymocytes from Twist2-transgenic mice were ChIPed with control IgG or anti-Twist2 antibody. A validated pair of primers for quantitative real-time PCR was used to amplify putative Twist2-binding sites within ROR $\gamma$ t promoter (indicated in Figure 24 as an asterisk). The data are presented as relative recovery rate for the signal from each sample relative to the chromatin lysates (Input). The negative control, Cyclophilin A (*CypA*) loci, showed no enrichment.



## DISCUSSION

T lymphocytes arise from hematopoietic stem cells through a series of tightly regulated processes. During development in the thymus, accessible fates of developing cells are progressively limited while being armed with effector gene expression (Rothenberg, 2011; Ramirez et al., 2010; Luc et al., 2012). Development and differentiation are characterized by the progressive changes in their cellular state which are mediated by the *de novo* binding of transcriptional factors to their cognate sites in genome (Britten and Davidson, 1969; McKenna and O'Malley, 2002). Recently, powerful technologies for investigating the interaction between transcription factors and DNA have been devised and a wealth of new insights has emerged. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) enables locating genomic regions, where critical regulatory interactions are taking place, in a high resolution (Barski et al., 2007; Johnson et al., 2007).

Twist2, a bHLH transcription factor, was reported to be induced in response to TCR signaling during thymocyte selection (Oh, 2016). Moreover, the expression of Twist2 is confined to the window of TCR signal strength that results in positive selection. Despite the importance of the positive selection-specific regulation of Twist2 expression for the proper differentiation of T lymphocytes, our understanding has been limited. In



this study, I aimed to characterize the biological role of Twist2 during positive selection by identifying its target genes. Through Twist2 ChIP-seq and further validation studies, I found out that Twist2 binds to the genomic loci pre-occupied by E proteins and thereby represses E protein target genes during positive selection.

*Genome-wide Twist2-binding sites were identified by ChIP-seq*

To investigate the genes regulated by Twist2, I performed ChIP-seq with thymocytes from the Twist2-transgenic mice and identified a total of 7,260 significant Twist2 ChIP-seq peaks. The mechanism how a transcription factor recognizes a limited set of targets from a lot of potential genomic binding sites has not been elucidated thoroughly. However, it is generally thought that genomic binding by transcription factor is guided to the open regions in chromatin and cell-selective occupancy of a given transcription factor is predetermined by chromosomal accessibility patterns (John et al., 2011). About half of the Twist2 ChIP-seq peaks (3,822/7,260) were located at the DNaseI hypersensitive sites in DP thymocyte genome (Vanhille et al., 2015), and the locations of these peaks were marked as highly confident Twist2-binding sites. The remaining 3,438 peaks were considered as the opportunistic genomic binding that may not translate into transcriptional regulation of neighboring genes. Indeed, the opportunistic genomic binding peaks are frequently observed in ChIP-seq analyses that were performed using

transgenic expression system (John et al., 2011; Zu et al., 2013).

Interestingly, gene-proximal localization of Twist2-binding sites was revealed when I analyzed the genome-wide distribution of Twist2-binding sites relative to the position of known and predicted genes. This preference of binding to proximal regions from transcription start sites seems to be unusual, since most ChIP-seq studies show the wider distribution of transcription factor binding across the entire genome. In addition, Twist2-binding sites appears to be evolutionarily conserved. The presence of strong evolutionary conservation over Twist2-binding sites suggests that these regions could be under selective pressure and therefore may be functionally important.

*E-box motifs were most significantly enriched in Twist2-binding sites*

The canonical E-box sequence, and two additional consensus sequences for the binding of ETS and Runx family transcription factors were found to be prevalent in Twist2-binding sites. This result of motif analysis does not seem to be only the case for Twist2-binding sites, since E proteins were also reported to bind to Ets or Runx motif-enriched sites in B- and T-cell precursors (Lin et al., 2010). Intriguingly, many ChIP-seq studies performed using hematopoietic lineage cells have repeatedly identified genomic regions that displayed cluster in binding of bHLH, Ets, and Runx family factors. Ets and Runx family factors are expressed throughout hematopoietic lineage development and

play significant roles in multipotent hematopoietic progenitors. In fact, most of hematopoietic cells, including T lymphocytes, preferentially utilize enhancers that are abundant in sites for the binding of Ets and Runx family transcription factors (Rothenberg, 2014). For example, the most active enhancers for the expression of signature genes in Th1 and Th2 cells are enriched for Runx motifs (Vahedi et al., 2012).

However, motif analysis results still suggest that Twist2 might interact with Ets or Runx family factors to regulate its target gene expression. Twist2 was reported to inhibit transcriptional activity of Runx1 (Sharabi et al., 2008) and Runx2 (Bialek et al., 2004; Fulzele et al., 2010) via physical interaction. In addition, it was also reported that TWIST1 protein could directly interact with ETV4, a member of ETS family transcription factors (Zhang et al., 2010). It remains to be determined whether physical interactions between these factors are important for the positive selection specific regulation of gene expression by Twist2.

E box is the consensus DNA sequence to which bHLH transcription factors bind. In general, bHLH proteins dimerize preferentially with other members of the bHLH family to form hetero dimers. The protein structure of the bHLH domain revealed that each subunit of the dimer adheres to one-half of the E-box sites (Ellenberger et al., 1994). A main factor, which influences a dimer choice, is the availability of bHLH proteins within cell. Before thymocyte selection, E protein transcription factors, HEB and E2A, are highly expressed in DP thymocytes and regulate the expression of their target genes

by binding to E-box sequences located in the regulatory regions. In particular, Twist2 was first identified from a yeast-two-hybrid screen using E2A as bait (Li et al., 1995). Based on these clues, I hypothesized that Twist2 may bind to the E-box sites pre-occupied by E proteins after being induced in post-DP thymocytes.

*Twist2 bind to overlapping sites with HEB on the DP-thymocytes genome and negatively regulates HEB target genes during thymocyte selection*

To validate this tentative hypothesis, a total of 1,655 HEB-binding sites in DP thymocytes (Vanhille et al., 2015) were analyzed and compared with Twist2-binding sites. Strikingly, the binding profiles of HEB and Twist2 displayed high degree of overlaps. 51% (842/1,655) of HEB-binding sites were intersected with Twist2-binding sites. Out of 430 putative genes directly regulated by E proteins in DP thymocytes, 239 genes were associated with the composite sites, binding both of Twist2 and HEB.

To further evaluate how the Twist2 occupancy correlates with the differential gene expression in DP thymocytes before (low level expression of Twist2) and immediately after (high level expression of Twist2) thymocyte selection, I analyzed the global gene expression profiles of pre-DP and post-DP thymocytes using microarray. In accordance with the hypothesis, those 239 genes that were associated with the composite sites were highly enriched among the group of genes that is down regulated upon

thymocyte selection. This result strongly reflected the functional consequences of Twist2 binding to the half of the sites formerly bound by E proteins in that Twist2 might negatively regulate E protein target genes during positive selection. Convincingly, these genes were also highly enriched among the group of genes down regulated upon E protein deletion, indicating that E-protein activity is indeed crucial for the proper expression in pre-DP thymocytes.

#### *Role of Twist2 in regulation of gene expression during thymocyte selection*

Out of the 239 genes associated with the composite sites, 73 genes showed significant differential expression during thymocyte selection. Among them, 23 genes were also differentially expressed upon deletion of E proteins. By using DO11.10 TCR-transgenic mouse system, I analyzed the effects of Twist2 knock out on the expression of these 23 genes. In the absence of Twist2, genes that are known to be down regulated in response to thymocyte selection were up regulated. For example, *Rorc* and *Cldn4*, whose expressions dropped dramatically upon receipt of positive selection signal, were up regulated in Twist2<sup>f/f</sup>CD4Cre<sup>+</sup> DO11.10 cells. In contrast, *Poll* and *Trim26*, whose expressions were highly induced in positively selected DP cells, were down regulated in Twist2<sup>f/f</sup>CD4Cre<sup>+</sup> DO11.10 cells.

Gene expression analysis displayed that some of Twist2 target genes were

activated, whereas others were repressed upon deletion of Twist2. This suggests bifunctional roles of Twist2 as activator and repressor of target gene expression. Twist1, another member of the Twist subfamily of bHLH proteins, has been known to play the bifunctional roles in regulating gene expression (reviewed in Franco et al., 2011). The last 20 amino acids at the C-terminus of Twist1, termed ‘Twist box’, contain a repressor domain (Bialek et al., 2004). A transactivation domain has also been characterized within the Twist box (Laursen et al., 2007) and the amino-acid residues Leu-187, Phe-191 and Arg-195 (LX3FX3R) are essential for the transactivation. Estimated modes of action comprise direct or indirect mechanisms: recruitment of coactivators or repressors after binding to E-box sites; inhibition of activator/repressor function through protein–protein interactions (reviewed in Franco et al., 2011). Twist2 exhibits a high degree of sequence similarity with Twist1. Furthermore, the Twist box is identical in both Twist1 and Twist2, suggesting that their functions might be redundant. Recently, the bifunctional roles of Twist2 have been reported in myeloid lineage development (Sharabi et al., 2008).

Nonetheless, Twist2 seems to act mainly as a repressor for the transcriptional regulation of E protein-mediated pre-selection programs in DP thymocytes, since the genes associated with the composite sites were highly enriched among the group of genes down regulated upon thymocyte selection. Further global gene expression analysis on the functional consequences of Twist2 knock out in DP thymocytes will clarify this issue thoroughly.

*Synergism between Twist2 and Id3 for the negative regulation of E protein target gene expression*

E protein transcription factors, HEB and E2A, are highly expressed at the DP stage and function as gatekeepers to block further development until a proper TCR is received (Jones and Zhuang, Immunity, 2007). As E-protein activity is down regulated by TCR-mediated positive selection signal, the DP to SP transition is initiated.

Up to date, Inhibitor of DNA binding 3 (Id3) is known as a sole factor for the down regulation of E-protein activity (Bain et al., 2001). The expression of Id3 is rapidly induced by TCR signal. Id3 protein bears an HLH domain, which lacks a basic DNA binding region. Therefore, ID3 can act as a negative regulator of E-protein activity via competitive dimerization with E proteins to inhibit DNA binding (Benezra et al., 1990). In the absence of Id3, positive selection is severely inhibited (Rivera et al., 2000). It is worth noting that thymocyte maturation is not completely blocked in the Id3-deficient mice. This raises a question about the existence of another unknown factors, which influence E-protein activity.

Here, I suggest Twist2 as a novel factor for the transcriptional repression of E protein-mediated pre-selection programs in DP thymocytes. After being highly induced in response to positively selecting TCR signals, Twist2 forms heterodimer with E proteins and thereby negatively regulates the expression of E protein target genes. As evidenced

by more severe inhibition of positive selection in the Twist2 and Id3 double knockout mice, it is conceivable that Twist2 and Id3 play redundant and synergistic roles for the progression of positive selection *in vivo* (unpublished data).

Although both Twist2 and Id3 dimerize with E proteins and negatively regulate E-protein activity, the mechanism of action seems to be different. In contrast to Id3, Twist2 may bind to E box as it dimerizes with E proteins and exert the negative regulation on the expression of neighboring genes. It will be interesting to determine a set of E protein target genes regulated uniquely by Twist2. ChIP-seq analyses of Twist2 and E proteins with the purified post-DP thymocytes will clarify this issue.

It is also intriguing that each Twist2 and Id3 are controlled by different signaling pathways, while they both are up regulated immediately after positive selection (Figure 30). Signals emanating from the TCR complex diverge into two downstream pathways. One is the mitogen-activated protein kinase (MAPK) pathway, and the other is the  $\text{Ca}^{2+}$ -Calcineurin pathway (Wang et al., 2010). The expression of Id3 is triggered by the MAPK pathway (Bain et al., 2001). However, Twist2 expression is under the control of NFATc3 that is activated by  $\text{Ca}^{2+}$ -Calcineurin pathway (Oh, 2010). Through various mouse models, it has well been studied that both ERK and Calcineurin are activated by TCR signaling that initiate the positive selection of thymocytes, whereas JNK is only activated in response to negative selection-strength of TCR (Gong et al., 2001). Consistently, ERK and Calcineurin-NFATc3 pathways support positive selection (Bueno et al., 2002; Cante-

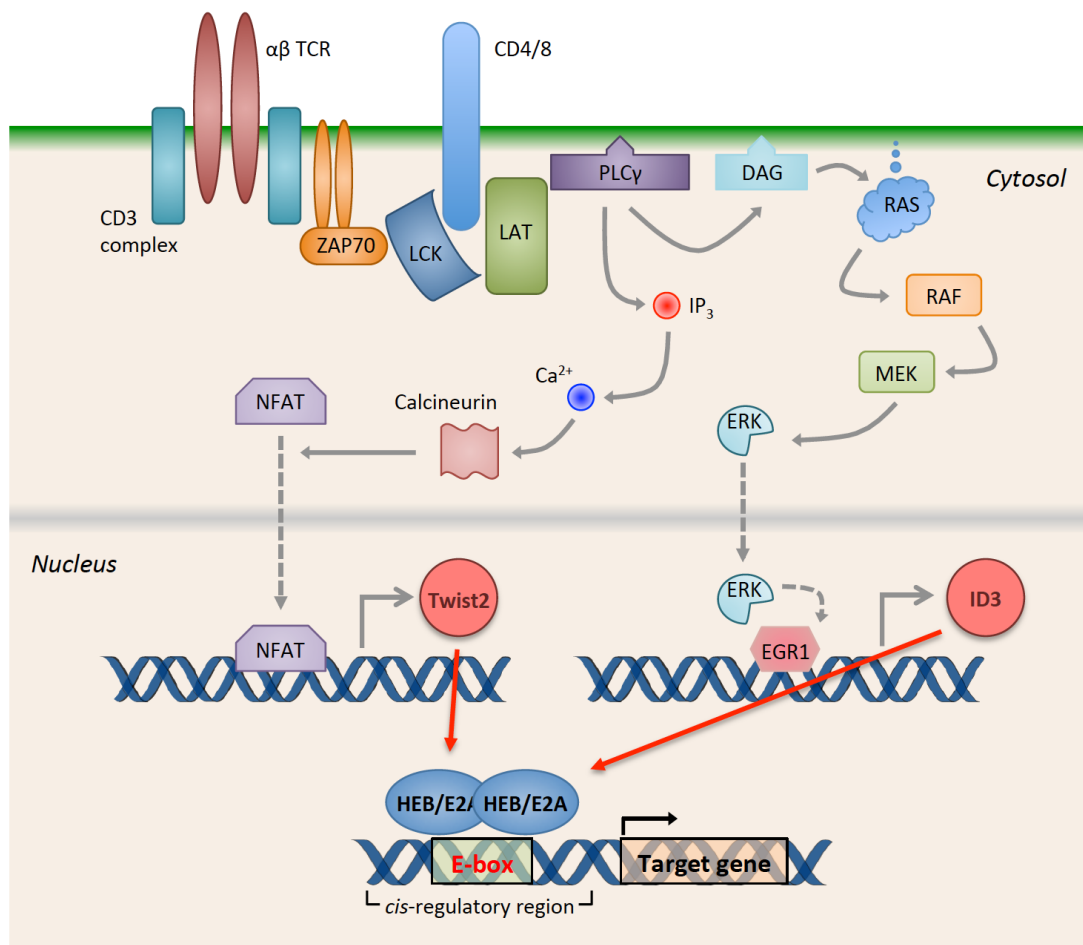


Barrett et al., 2007; Fischer et al., 2005), while the JNK pathway is essential for negative selection (Rincon et al., 1998; Sabapathy et al., 2001).

In particular, the expression of Twist2 ceases in thymocytes undergoing negative selection as NFATc3 is exported to cytoplasm by JNK. Considering that Twist2 binds to about half of E protein target genes, there may be different gene expression programs between positive selection and negative selection of thymocytes, which are guided by differential expression of Twist2.

### *Conclusion*

Recent studies have identified bHLH proteins as important regulators of T cell development. My study provides the evidences that Twist2 functions to regulate the expression of E protein target genes during thymocyte selection (Figure 29). I want to highlight that Twist2 is induced through  $\text{Ca}^{2+}$ -Calcineurin pathway. It will be interesting to examine whether Twist2 plays critical role as an integrator interpreting TCR signals transmitted through  $\text{Ca}^{2+}$ -Calcineurin pathway.



**Figure 29. Model of the Twist2-mediated regulation of gene expression during thymocyte selection**

## REFERENCES

1. Agata, Y. et al. Regulation of T cell receptor beta gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity* 27, 871–884 (2007).
2. Aifantis, I., Feinberg, J., Fehling, H. J., Di Santo, J. P. & von Boehmer, H. Early T cell receptor beta gene expression is regulated by the pre-T cell receptor-CD3 complex. *J. Exp. Med.* 190, 141–144 (1999).
3. Alberola-Ila, J., Forbush, K. A., Seger, R., Krebs, E. G. & Perlmutter, R. M. Selective requirement for MAP kinase activation in thymocyte differentiation. *Nature* 373, 620–623 (1995).
4. Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T. & Saltiel, A. R. PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase in vitro and in vivo. *J. Biol. Chem.* 270, 27489–27494 (1995).
5. Bain, G. et al. Regulation of the helix-loop-helix proteins, E2A and Id3, by the Ras-ERK MAPK cascade. *Nat. Immunol.* 2, 165–171 (2001).
6. Bain, G. et al. E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. *Mol. Cell. Biol.* 17, 4782–4791 (1997).
7. Bain, G. et al. E2A proteins are required for proper B cell development and initiation

- of immunoglobulin gene rearrangements. *Cell* 79, 885–892 (1994).
8. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837 (2007).
9. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weintraub, H. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 61, 49–59 (1990).
10. Bialek, P. et al. A twist code determines the onset of osteoblast differentiation. *Dev. Cell* 6, 423–435 (2004).
11. Biddison, W. E., Rao, P. E., Talle, M. A., Goldstein, G. & Shaw, S. Possible involvement of the OKT4 molecule in T cell recognition of class II HLA antigens. Evidence from studies of cytotoxic T lymphocytes specific for SB antigens. *J. Exp. Med.* 156, 1065–1076 (1982).
12. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322 (2008).
13. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* 165, 349–357 (1969).
14. Cabaniols, J. P., Fazilleau, N., Casrouge, A., Kourilsky, P. & Kanellopoulos, J. M. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J. Exp. Med.* 194, 1385–1390 (2001).

15. Canté-Barrett, K., Winslow, M. M. & Crabtree, G. R. Selective role of NFATc3 in positive selection of thymocytes. *J. Immunol.* 179, 103–110 (2007).
16. Chothia, C., Boswell, D. R. & Lesk, A. M. The outline structure of the T-cell alpha beta receptor. *EMBO J.* 7, 3745–3755 (1988).
17. David-Fung, E.-S. et al. Progression of regulatory gene expression states in fetal and adult pro-T-cell development. *Immunol. Rev.* 209, 212–236 (2006).
18. Dower, N. A. et al. RasGRP is essential for mouse thymocyte differentiation and TCR signaling. *Nat. Immunol.* 1, 317–321 (2000).
19. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24, 1547–1548 (2008).
20. Dudley, D. T., Pang, L., Decker, S. J., Bridges, A. J. & Saltiel, A. R. A synthetic inhibitor of the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7686–7689 (1995).
21. Eberl, G. et al. An essential function for the nuclear receptor RORgamma(t) in the generation of fetal lymphoid tissue inducer cells. *Nat. Immunol.* 5, 64–73 (2004).
22. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S. C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8, 970–980 (1994).

23. Engel, I., Johns, C., Bain, G., Rivera, R. R. & Murre, C. Early thymocyte development is regulated by modulation of E2A protein activity. *J. Exp. Med.* 194, 733–745 (2001).
24. Engel, I. & Murre, C. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO J.* 23, 202–211 (2004).
25. Feske, S. Calcium signalling in lymphocyte activation and disease. *Nat. Rev. Immunol.* 7, 690–702 (2007).
26. Fischer, A. M., Katayama, C. D., Pagès, G., Pouyssegur, J. & Hedrick, S. M. The role of erk1 and erk2 in multiple stages of T cell development. *Immunity* 23, 431–443 (2005).
27. Franco, H. L., Casasnovas, J., Rodríguez-Medina, J. R. & Cadilla, C. L. Redundant or separate entities? - roles of Twist1 and Twist2 as molecular switches during gene transcription. *Nucleic Acids Res.* 39, 1177–1186 (2011).
28. Freeman, M. Feedback control of intercellular signalling in development. *Nature* 408, 313–319 (2000).
29. Fulzele, K. et al. Insulin receptor signaling in osteoblasts regulates postnatal bone acquisition and body composition. *Cell* 142, 309–319 (2010).
30. Gao, E. K., Lo, D., Cheney, R., Kanagawa, O. & Sprent, J. Abnormal differentiation of thymocytes in mice treated with cyclosporin A. *Nature* 336, 176–179 (1988).
31. Germain, R. N. MHC-dependent antigen processing and peptide presentation:

providing ligands for T lymphocyte activation. *Cell* 76, 287–299 (1994).

32. Godfrey, D. I., Kennedy, J., Suda, T. & Zlotnik, A. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *J. Immunol.* 150, 4244–4252 (1993).

33. Groettrup, M. et al. A novel disulfide-linked heterodimer on pre-T cells consists of the T cell receptor beta chain and a 33 kd glycoprotein. *Cell* 75, 283–294 (1993).

34. Hashimoto, K. et al. Requirement for p56lck tyrosine kinase activation in T cell receptor-mediated thymic selection. *J. Exp. Med.* 184, 931–943 (1996).

35. He, Y.-W. Orphan nuclear receptors in T lymphocyte development. *J. Leukoc. Biol.* 72, 440–446 (2002).

36. He, Y. W. et al. Down-regulation of the orphan nuclear receptor ROR gamma t is essential for T lymphocyte maturation. *J. Immunol.* 164, 5668–5674 (2000).

37. He, Y. W., Deftos, M. L., Ojala, E. W. & Bevan, M. J. RORgamma t, a novel isoform of an orphan receptor, negatively regulates Fas ligand expression and IL-2 production in T cells. *Immunity* 9, 797–806 (1998).

38. Henthorn, P., Kiledjian, M. & Kadesch, T. Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. *Science* 247, 467–470 (1990).

39. Homer, N., Merriman, B. & Nelson, S. F. BFAST: An Alignment Tool for Large Scale Genome Resequencing. *PLoS ONE* 4, e7767 (2009).
40. Hu, J. S., Olson, E. N. & Kingston, R. E. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Mol. Cell. Biol.* 12, 1031–1042 (1992).
41. Janeway, C. A. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu. Rev. Immunol.* 10, 645–674 (1992).
42. Jenkins, M. K., Schwartz, R. H. & Pardoll, D. M. Effects of cyclosporine A on T cell development and clonal deletion. *Science* 241, 1655–1658 (1988).
43. Jetten, A. M., Kurebayashi, S. & Ueda, E. The ROR nuclear orphan receptor subfamily: critical regulators of multiple biological processes. *Prog. Nucleic Acid Res. Mol. Biol.* 69, 205–247 (2001).
44. Jetten, A. M. & Ueda, E. Retinoid-related orphan receptors (RORs): roles in cell survival, differentiation and disease. *Cell Death Differ.* 9, 1167–1171 (2002).
45. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 43, 264–268 (2011).
46. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of



in vivo protein-DNA interactions. *Science* 316, 1497–1502 (2007).

47. Jones, M. E. & Zhuang, Y. Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. *Immunity* 27, 860–870 (2007).

48. Koltsova, E. K. et al. Early growth response 1 and NF-ATc1 act in concert to promote thymocyte development beyond the beta-selection checkpoint. *J. Immunol.* 179, 4694–4703 (2007).

49. Kurebayashi, S. et al. Retinoid-related orphan receptor gamma (RORgamma) is essential for lymphoid organogenesis and controls apoptosis during thymopoiesis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10132–10137 (2000).

50. Laursen, K. B., Mielke, E., Iannaccone, P. & Füchtbauer, E.-M. Mechanism of transcriptional activation by the proto-oncogene Twist1. *J. Biol. Chem.* 282, 34623–34633 (2007).

51. Lee, M. S., Lowe, G., Flanagan, S., Kuchler, K. & Glackin, C. A. Human Dermo-1 has attributes similar to twist in early bone development. *Bone* 27, 591–602 (2000).

52. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 1, 729–748 (2006).

53. Lepoivre, C. et al. Divergent transcription is associated with promoters of

transcriptional regulators. *BMC Genomics* 14, 914 (2013).

54. Li, L., Cserjesi, P. & Olson, E. N. Dermo-1: a novel twist-related bHLH protein expressed in the developing dermis. *Dev. Biol.* 172, 280–292 (1995).

55. Lin, Y. C. et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* 11, 635–643 (2010).

56. Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biology* 12, R83 (2011).

57. Luc, S. et al. The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat. Immunol.* 13, 412–419 (2012).

58. Lundberg, K., Heath, W., Köntgen, F., Carbone, F. R. & Shortman, K. Intermediate steps in positive selection: differentiation of CD4<sup>+</sup>8<sup>int</sup> TCR<sup>int</sup> thymocytes into CD4<sup>+</sup>8<sup>+</sup>TCR<sup>hi</sup> thymocytes. *J. Exp. Med.* 181, 1643–1651 (1995).

59. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697 (2011).

60. Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* 20, 429–440 (2000).

61. McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* 108, 465–474 (2002).

62. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotech* 28, 495–501 (2010).
63. Medvedev, A., Chistokhina, A., Hirose, T. & Jetten, A. M. Genomic structure and chromosomal mapping of the nuclear orphan receptor ROR gamma (RORC) gene. *Genomics* 46, 93–102 (1997).
64. Mellentin, J. D. et al. The gene for enhancer binding proteins E12/E47 lies at the t(1;19) breakpoint in acute leukemias. *Science* 246, 379–382 (1989).
65. Michie, A. M. et al. Clonal characterization of a bipotent T cell and NK cell progenitor in the mouse fetal thymus. *J. Immunol.* 164, 1730–1733 (2000).
66. Mombaerts, P. et al. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* 68, 869–877 (1992).
67. Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56, 777–783 (1989).
68. Murre, C. et al. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58, 537–544 (1989).
69. Murre, C. Helix-loop-helix proteins and lymphocyte development. *Nat. Immunol.* 6, 1079–1086 (2005).

70. Negishi, I. et al. Essential role for ZAP-70 in both positive and negative selection of thymocytes. *Nature* 376, 435–438 (1995).
71. Neilson, J. R., Winslow, M. M., Hur, E. M. & Crabtree, G. R. Calcineurin B1 is essential for positive but not negative selection during thymocyte development. *Immunity* 20, 255–266 (2004).
72. O'Shea, C. C., Crompton, T., Rosewell, I. R., Hayday, A. C. & Owen, M. J. Raf regulates positive selection. *Eur. J. Immunol.* 26, 2350–2355 (1996).
73. Oukka, M. et al. The transcription factor NFAT4 is involved in the generation and survival of T cells. *Immunity* 9, 295–304 (1998).
74. Pagès, G. et al. Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* 286, 1374–1377 (1999).
75. Pan, L., Hanrahan, J., Li, J., Hale, L. P. & Zhuang, Y. An analysis of T cell intrinsic roles of E2A by conditional gene disruption in the thymus. *J. Immunol.* 168, 3923–3932 (2002).
76. Pannetier, C. et al. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4319–4323 (1993).
77. Pui, J. C. et al. Notch1 expression in early lymphopoiesis influences B versus T

lineage determination. *Immunity* 11, 299–308 (1999).

78. Radtke, F. et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* 10, 547–558 (1999).

79. Ramírez, J., Lukin, K. & Hagman, J. From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Curr. Opin. Immunol.* 22, 177–184 (2010).

80. Rao, A., Luo, C. & Hogan, P. G. Transcription factors of the NFAT family: regulation and function. *Annu. Rev. Immunol.* 15, 707–747 (1997).

81. Rincón, M., Flavell, R. A. & Davis, R. A. The JNK and P38 MAP kinase signaling pathways in T cell-mediated immune responses. *Free Radic. Biol. Med.* 28, 1328–1337 (2000).

82. Rivera, R. R., Johns, C. P., Quan, J., Johnson, R. S. & Murre, C. Thymocyte selection is regulated by the helix-loop-helix inhibitor protein, Id3. *Immunity* 12, 17–26 (2000).

83. Robey, E. & Fowlkes, B. J. Selective events in T cell development. *Annu. Rev. Immunol.* 12, 675–705 (1994).

84. Robinson, J. T. et al. Integrative genomics viewer. *Nat Biotech* 29, 24–26 (2011).

85. Rock, E. P., Sibbald, P. R., Davis, M. M. & Chien, Y. H. CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* 179, 323–328 (1994).

86. Rothenberg, E. V. T cell lineage commitment: identity and renunciation. *J. Immunol.* 186, 6649–6655 (2011).
87. Sharabi, A. B. et al. Twist-2 controls myeloid lineage development and function. *PLoS Biol.* 6, e316 (2008).
88. Shinkai, Y. et al. Restoration of T cell development in RAG-2-deficient mice by functional TCR transgenes. *Science* 259, 822–825 (1993).
89. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050 (2005).
90. Sohn, S. J., Forbush, K. A., Pan, X. C. & Perlmutter, R. M. Activated p56lck directs maturation of both CD4 and CD8 single-positive thymocytes. *J. Immunol.* 166, 2209–2217 (2001).
91. Šošić, D., Richardson, J. A., Yu, K., Ornitz, D. M. & Olson, E. N. Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. *Cell* 112, 169–180 (2003).
92. Sun, Z. et al. Requirement for RORgamma in thymocyte survival and lymphoid organ development. *Science* 288, 2369–2373 (2000).
93. Swain, S. L. T cell subsets and the recognition of MHC class. *Immunol. Rev.* 74, 129–142 (1983).

94. Swan, K. A. et al. Involvement of p21ras distinguishes positive and negative selection in thymocytes. *EMBO J.* 14, 276–285 (1995).
95. Tükel, T. et al. Homozygous nonsense mutations in TWIST2 cause Setleis syndrome. *Am. J. Hum. Genet.* 87, 289–296 (2010).
96. Vahedi, G. et al. STATs shape the active enhancer landscape of T cell populations. *Cell* 151, 981–993 (2012).
97. Vanhille, L. et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6, 6905 (2015).
98. van Oers, N. S., von Boehmer, H. & Weiss, A. The pre-T cell receptor (TCR) complex is functionally coupled to the TCR-zeta subunit. *J. Exp. Med.* 182, 1585–1590 (1995).
99. Villey, I., de Chasseval, R. & de Villartay, J. P. RORgammaT, a thymus-specific isoform of the orphan nuclear receptor RORgamma / TOR, is up-regulated by signaling through the pre-T cell receptor and binds to the TEA promoter. *Eur. J. Immunol.* 29, 4072–4080 (1999).
100. von Boehmer, H. & Fehling, H. J. Structure and function of the pre-T cell receptor. *Annu. Rev. Immunol.* 15, 433–452 (1997).
101. von Boehmer, H., Teh, H. S. & Kisielow, P. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol. Today* 10, 57–61 (1989).

102. Wang, D. et al. The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. *J. Immunol.* 177, 109–119 (2006).
103. Wange, R. L. LAT, the linker for activation of T cells: a bridge between T cell-specific and general signaling pathways. *Sci. STKE* 2000, re1 (2000).
104. Wilson, A., MacDonald, H. R. & Radtke, F. Notch 1-deficient common lymphoid precursors adopt a B cell fate in the thymus. *J. Exp. Med.* 194, 1003–1012 (2001).
105. Wilson-Rawls, J., Rhee, J. M. & Rawls, A. Paraxis is a basic helix-loop-helix protein that positively regulates transcription through binding to specific E-box elements. *J. Biol. Chem.* 279, 37685–37692 (2004).
106. Wolf, C. et al. The M-twist gene of *Mus* is expressed in subsets of mesodermal cells and is closely related to the *Xenopus* X-twi and the *Drosophila* twist genes. *Dev. Biol.* 143, 363–373 (1991).
107. Xi, H., Schwartz, R., Engel, I., Murre, C. & Kersh, G. J. Interplay between RORgammat, Egr3, and E proteins controls proliferation in response to pre-TCR signals. *Immunity* 24, 813–826 (2006).
108. Zhang, W. et al. Essential role of LAT in T cell development. *Immunity* 10, 323–332 (1999).



109. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).
110. Zhang, Z. et al. Preaxial polydactyly: interactions among ETV, TWIST1 and HAND2 control anterior-posterior patterning of the limb. *Development* 137, 3417–3426 (2010).
111. Zhu, B.-M. et al. Genome-wide analyses reveal the extent of opportunistic STAT5 binding that does not yield transcriptional activation of neighboring genes. *Nucleic Acids Res.* 40, 4461–4472 (2012).

## ABSTRACT IN KOREAN

흉선 세포 선별과정에서 E 단백질의 타겟 유전자 발현을 억제하는

Twist2 의 기능에 관한 연구

오 승 준

T 세포는 골수에서 발생한 전구 세포로부터 유래하며, 흉선 내에서 복잡한 선별 과정을 거쳐 생성된다. 특별히 T 세포 수용체(TCR) 유전자의 재조합을 마친 DP 단계에서 긍정적 선별(positive selection)이 일어나게 되는데, 이를 통해 주조직적합성항원(MHC)과 펩티드 항원을 적절한 강도로 인식하는 세포들만이 세포 소멸에서 구제되어 성숙한 T 세포로 분화하게 된다. 긍정적 선별은 수용체와 같은 세포 표면 단백질과 세포 내 신호 전달 물질, 그리고 전사 조절 단백질 등의 다양한 분자들에 의하여 까다롭게 조절된다.

본 논문에서 본인은 흉선 세포의 긍정적 선별 과정에서 발현되는 전사

조절 인자인 Twist2가 수행하는 역할에 대한 연구를 수행하였다. 이 연구를 통해, 긍정적 선별 신호에 의해 발현이 유도된 Twist2가 DP 단계의 유전자 발현을 총괄하는 E 단백질(E protein)을 억제하는 기능을 수행함을 새롭게 밝혀내었다. 구체적으로, 최신 유전체 분석 기술인 ChIP-seq을 활용해 Twist2가 E 단백질에 의해 점유되는 유전자 부위에 결합할 수 있음을 보여주었다. 그리고 여기에서 얻어진 정보를 바탕으로 전사체 분석과 생화학적 검증 실험을 수행하여 긍정적 선별 과정 동안 Twist2에 의해 타겟 유전자들의 발현이 감소되는 것을 확인하였다.

주요어 : Twist2, 흥선세포, 긍정적선별, ChIP-seq, 전사체분석

학번 : 2008-30808